

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:38:48 ; Search time 9.06061 Seconds
(without alignments)
1368.719 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1647	100.0	299	1 MTL5_HUMAN	Q9Y415 homo sapien
2	1251	76.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
3	109.5	6.6	769	1 ITB2_HUMAN	P05107 homo sapien
4	106.5	6.5	3718	1 LMA5_MOUSE	Q61001 mus musculus
5	103.5	6.3	1786	1 LMB1_MOUSE	P02469 mus musculus
6	103	6.3	769	1 ITB2_PIG	P53714 sus scrofa
7	102	6.2	1790	1 LMB1_DROME	P11046 drosophila
8	101.5	6.2	396	1 TFH2_MOUSE	Q9j1b4 mus musculus
9	101	6.1	1700	1 BAR3_CHITE	Q03376 chironomus
10	101	6.1	3011	1 POLG_HCV1	P26664 h genome po
11	100.5	6.1	760	1 E2_DROME	P42124 drosophila
12	100	6.1	3595	1 LMA5_HUMAN	O15230 homo sapien
13	99	6.0	787	1 ITB6_MOUSE	Q920t9 mus musculus
14	98.5	6.0	3106	1 LMA2_MOUSE	Q60675 mus musculus
15	98.5	6.0	5376	1 ZAN_MOUSE	O88799 mus musculus
16	97.5	5.9	328	1 C170_GIALA	P15799 giardia lam
17	97.5	5.9	643	1 CD93_RAT	Q9et61 rattus norv
18	97.5	5.9	969	1 PAC4_HUMAN	P29122 homo sapien
19	97.5	5.9	971	1 RECK_HUMAN	O95980 homo sapien
20	97	5.9	395	1 TFH2_HUMAN	P13888 homo sapien
21	97	5.9	798	1 ITB7_HUMAN	P26010 homo sapien
22	97	5.9	3010	1 POLG_HCVJT	Q00269 h genome po
23	96.5	5.9	290	1 Y070_NPVAC	P41470 autographa
24	96.5	5.9	1576	1 YLK3_CABEL	P41951 caenorhabdi
25	96.5	5.9	3726	1 TRX_DROME	P20659 drosophila
26	96	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
27	96	5.8	1808	1 TENA_CHICK	P10039 gallus gall
28	96	5.8	3010	1 POLG_HCVBK	P26663 h genome po
29	95.5	5.8	747	1 E2H1_HUMAN	Q92800 homo sapien
30	95.5	5.8	1801	1 LMB2_RAT	P15800 rattus norv
31	94	5.7	3011	1 POLG_HCVH	P27958 h genome po
32	93	5.6	3110	1 LMA2_HUMAN	P24043 homo sapien
33	93	5.6	3224	1 RBP2_HUMAN	P49792 homo sapien

34	92.5	5.6	332	1 XYNA_CRYAL	P07529 cryptococu
35	92.5	5.6	611	1 LEM2_CANFA	P33730 canis famil
36	92.5	5.6	773	1 MSU2_DROME	P50534 drosophila
37	92.5	5.6	1172	1 AHM2_ARATH	O64474 arabidopsis
38	92	5.6	798	1 ITB0_XENLA	P12607 xenopus lae
39	92	5.6	3075	1 LMA1_HUMAN	P25391 homo sapien
40	91.5	5.6	4544	1 LRP1_HUMAN	Q07954 homo sapien
41	90.5	5.5	354	1 VEGD_HUMAN	O43915 homo sapien
42	90.5	5.5	747	1 E2H1_MOUSE	P70351 mus musculus
43	90.5	5.5	1786	1 LMB1_HUMAN	P07942 homo sapien
44	90	5.5	3010	1 POLG_HCVTW	P29846 h genome po
45	89.5	5.4	771	1 ITB2_MOUSE	P11835 mus musculus

ALIGNMENTS

RESULT 1

ID	MTL5_HUMAN	STANDARD;	PRT;	299 AA.
AC	Q9Y415;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific			
DE	metallothionein-like protein).			
GN	MTL5			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=99208669; PubMed=10191092;			
RA	Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;			
RT	"A novel testis-specific metallothionein-like protein, tesmin, is an			
RT	early marker of male germ cell differentiation.";			
RL	Genomics 57:130-136(1999).			
CC	FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.			
CC	TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collabora-			
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; U86074; AAD24668.1;			
DR	Genew; HGNC:7446; MTL5.			
DR	MIM; 604374; -			
DR	InterPro; IPR005172; CXC.			
DR	Pfam; PF03638; CXC; 1.			
KW	Spermatogenesis.			
FT	DOMAIN 99 191			
FT	CYS-RICH			
SQ	SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;			
Query Match	100.0%;	Score 1647;	DB 1;	Length 299;
Best Local Similarity	100.0%;	Pred. No. 2.5e-134;		
Matches 299;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MVICQLKGGTQMLCIDNSRTRELKALHVPQYQDQNNYQSDVPKPMTALVGRFLPASTK	60	
Db	1	MVICQLKGGTQMLCIDNSRTRELKALHVPQYQDQNNYQSDVPKPMTALVGRFLPASTK	60	
QY	61	LNLTITQLEGALPSVVVNGSAFSGSTLPKPPKITTLAGYDCDFASGDFCNCCNCCNL	120	
Db	61	LNLTITQLEGALPSVVVNGSAFSGSTLPKPPKITTLAGYDCDFASGDFCNCCNCCNL	120	
QY	121	HHDIERFRAIKACLGRNPFAFPKIGKGLGNVYKPOHNGCNCRRSGCLKNYCEYEAI	180	
Db	121	HHDIERFRAIKACLGRNPFAFPKIGKGLGNVYKPOHNGCNCRRSGCLKNYCEYEAI	180	

```

Db 121 HHDIERFAKRAICLRNPEAFQPKIGKQGLGNVKNQHNKGCNRRSGCLKNYCEYEAQI 180
Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDHRRPSS 240
Db 181 MCSSICKICGCKNYEESPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDHRRPSS 240
Qy 241 CISWVVEATCACLAAQGEAEKHCSCKLAEDMTLEEFGRCLSOILHTEFKSKGLKME 299
Db 241 CISWVVEATCACLAAQGEAEKHCSCKLAEDMTLEEFGRCLSOILHTEFKSKGLKME 299

RESULT 2
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE tesmin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
GN MTL5
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -
DR EMBL; U67176; AAD24666.1; -
DR MGD; MGI:1340029; MTL5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 76.0%; Score 1251; DB 1; Length 295;
Best Local Similarity 76.3%; Pred. No. 2.5e-100;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICOLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMFALYGRFLPASTK 60
Db 1 MVICOLKGAQMCLCIDNGARELALHLLPQYDQSSFPQSELPKPMTTLVGRLLPVPK 60
Qy 61 LNLITQLEGALPVSNGSAPSPSTLPGLPKITLAGYCDGFCFSGDNCNCCNNL 120
Db 61 LNLITQVNGALPALSVAINGAAPSPGALQGPPLKITLSGYCDGFCFSGDNCSCS----CNL 116
Qy 121 HHDIERFAKRAICLRNPEAFQPKIGKQGLGNVKNQHNKGCNRRSGCLKNYCEYEAQI 180
Db 117 RHELFKRAICLRNPEAFQPKRMGRGLGAALKRSGKNCRRSGCLKNYCEYEAQI 176
Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

```

```

Db 177 MCSSICKICAKNYEESPERKMLSTHYMEPGDFESHYLSPAKFSPPKLRNRQAFS 236
Qy 241 CISWVVEATCACLAAQGEAEKHCSCKLAEDMTLEEFGRCLSOILHTEFKSKGLKME 299
Db 237 CISWVVEATCACLAAQGEAEKHCSPSLAEQMILEEFGRCLSOILHTEFKSKGLKIE 295

RESULT 3
ITB2_HUMAN
ID ITB2_HUMAN STANDARD; PRT; 769 AA.
AC P05107; Q16418;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/p150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
GN ITGB2 OR CD18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131080; PubMed=3028646;
RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
RT "Cloning of the beta subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family.";
RL Cell 48:681-690(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92077153; PubMed=1683838;
RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
RL FEBS Lett. 294:97-103(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P., Schafte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=87246525; PubMed=2954816;
RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C., Wong A.J.;
RT "The primary structure of the beta-subunit of the cell surface adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to the fibronectin receptor.";
RL EMBO J. 6:915-919(1987).
RN [5]
RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
RX MEDLINE=92147694; PubMed=1346613;
RA Nelson C., Rabb H., Arnaout M.A.;
RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing and a missense mutation in a conserved region of CD18 impair cell surface expression of beta 2 integrins.";
RL J. Biol. Chem. 267:3351-3357(1992).
RN [6]

```

RP VARIANTS LAD1 PRO-149 AND ARG-169.
 RX MEDLINE-90293696; PubMed-1694220;
 RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;
 RT "Distinct mutations in two patients with leukocyte adhesion
 deficiency and their functional correlates";
 RL J. Exp. Med. 172:335-345(1990).
 RN [7]
 RP VARIANT LAD1 ARG-169.
 RX MEDLINE-92324303; PubMed-1352501;
 RA Corbi A., Vara A., Ursa A., Rodriguez M.C.G., Fontan G.,
 RA Sanchez-Madrid F.;
 RT "Molecular basis for a severe case of leukocyte adhesion deficiency.";
 RL Eur. J. Immunol. 22:1877-1881(1992).
 RN [8]
 RP VARIANTS LAD1 THR-196 AND CYS-593.
 RX MEDLINE-90187099; PubMed-1968911;
 RA Arnout M.A., Dana N., Gupta S.K., Tenen D.G., Fathallah D.M.;
 RT "Point mutations impairing cell surface expression of the common beta
 subunit (CD18) in a patient with leukocyte adhesion molecule
 (Leu-CAM) deficiency.";
 RL J. Clin. Invest. 85:977-981(1990).
 RN [9]
 RP VARIANT LAD1 LEU-178.
 RX MEDLINE-92184805; PubMed-1347532;
 RA Back L.L., Kwok W.W., Hickstein D.D.;
 RT "Identification of two molecular defects in a child with leukocyte
 adherence deficiency.";
 RL J. Biol. Chem. 267:5482-5487(1992).
 RN [10]
 RP VARIANT LAD1 ASN-128.
 RX MEDLINE-92272746; PubMed-1590804;
 RA Matsuura S., Kishi F., Tsukahara M., Nunci H., Matsuda I.,
 RA Kobayashi K., Kajii T.;
 RT "Leukocyte adhesion deficiency: identification of novel mutations in
 two Japanese patients with a severe form.";
 RL Biochem. Biophys. Res. Commun. 184:1460-1467(1992).
 RN [11]
 RP VARIANT LAD1 SER-284.
 RX MEDLINE-93312347; PubMed-7686755;
 RA Back L.A., Kerkering M., Baker D., Bauer T.R., Embree L.J.,
 RA Hickstein D.D.;
 RT "A point mutation associated with leukocyte adhesion deficiency type
 1 of moderate severity";
 RL Biochem. Biophys. Res. Commun. 193:912-918(1993).
 RN [12]
 RP VARIANTS LAD1 PRO-138 AND ARG-273.
 RX MEDLINE-99102253; PubMed-9884339;
 RA Hogg N., Stewart M.P., Scarth S.L., Newton R., Shaw J.M., Law S.K.A.,
 RA Klein N.;
 RT "A novel leukocyte adhesion deficiency caused by expressed but
 nonfunctional beta2 integrins Mac-1 and LFA-1.";
 RL J. Clin. Invest. 103:97-106(1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 CC ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
 CC RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
 CC AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE
 CC SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
 CC RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
 CC ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
 CC D/BETA 2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
 CC -1- SUBUNIT: HETEROIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
 CC ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -1- DISEASE: DEFECTS IN ITGB2 ARE THE CAUSE OF LEUKOCYTE ADHESION
 CC DEFICIENCY, TYPE I (LAD1). LAD1 PATIENTS HAVE RECURRENT BACTERIAL
 CC INFECTIONS AND THEIR LEUKOCYTES ARE DEFICIENT IN A WIDE RANGE OF
 CC ADHESION-DEPENDENT FUNCTIONS.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD18 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd18.htm".

 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y00057; CAA68266.1;
 DR EMBL; X64072; CAA45427.1;
 DR EMBL; X64073; CAA45427.1; JOINED.
 DR EMBL; X64074; CAA45427.1; JOINED.
 DR EMBL; X64075; CAA45427.1; JOINED.
 DR EMBL; X64076; CAA45427.1; JOINED.
 DR EMBL; X64077; CAA45427.1; JOINED.
 DR EMBL; X64078; CAA45427.1; JOINED.
 DR EMBL; X64079; CAA45427.1; JOINED.
 DR EMBL; X64080; CAA45427.1; JOINED.
 DR EMBL; X64081; CAA45427.1; JOINED.
 DR EMBL; X64082; CAA45427.1; JOINED.
 DR EMBL; X64083; CAA45427.1; JOINED.
 DR EMBL; X63924; CAA45427.1; JOINED.
 DR EMBL; X63925; CAA45427.1; JOINED.
 DR EMBL; X63926; CAA45427.1; JOINED.
 DR EMBL; AL163300; CAB90553.1;
 DR EMBL; M15395; AAA59490.1;
 DR EMBL; S81234; AAB21404.1;
 DR PIR; A25967; LJHULM.
 DR PIR; A29265; A29265.
 DR HSSP; P05106; LJV2.
 DR GENE; HGNC:6155; ITGB2.
 DR MIM; 600065;
 DR MIM; 116920;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR ProDom; PD001811; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal; Disease mutation.
 FT SIGNAL 1 22
 FT CHAIN 23 769 INTEGRIN BETA-2.
 FT DOMAIN 23 700 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 701 723 POTENTIAL.
 FT DOMAIN 724 769 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 134 363 VWFA-LIKE.
 FT DOMAIN 449 617 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 449 496 I.
 FT REPEAT 497 540 II.
 FT REPEAT 541 581 III.
 FT REPEAT 582 617 IV.
 FT DISULFID 25 447 BY SIMILARITY.
 FT DISULFID 33 43 BY SIMILARITY.
 FT DISULFID 36 73 BY SIMILARITY.
 FT DISULFID 46 62 BY SIMILARITY.
 FT DISULFID 191 198 BY SIMILARITY.

Query Match 6.6%; Score 109.5; DB 1; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.069;
 Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
 QY 1 MWICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYLQSDVPKPMVTLVGRFLPASTK 60
 DB 436 IIVTVQLPQCECRCDQSRDRL--CH-----GKGFLECGICRCDTGYIGNKCECQTQ 486

FT	DISULFID	436	BY SIMILARITY.	452	QY	94	TLGYCD-CFA-----SCDFCNCNCNCNHLHDIERFKAIRKACLRNPE	139
FT	DISULFID	454	BY SIMILARITY.	463	DB	333	TCGGSCDRCCPGFNOOPWKPATYDSANECQSCNCHGAYDYDPEVDR-----RNAS	385
FT	DISULFID	466	BY SIMILARITY.	476	QY	140	AFQPKIGKQGLGNV---KPOHNKGCNCRSGCLKNYCEYEAQIMCSCICKIGCKN---	193
FT	DISULFID	500	BY SIMILARITY.	512	DB	386	QNQDNVYQG--GGVCLDCOHHITGINCER--CLPGFFRAPDQPLDSPHVCRPCDCSDFT	441
FT	DISULFID	521	BY SIMILARITY.	532	QY	194	---YEESPERKTLMSPNYMQTGL-----EG-----SHYLPPTKFGSLPRFESH-DRRPS	240
FT	DISULFID	523	BY SIMILARITY.	534	DB	442	DGTCEDLTGR--CYCRPNF--TGELCAACAEGYTDPPHCYP-----LPSFPHNDTREQV	491
FT	DISULFID	547	BY SIMILARITY.	559	QY	241	CISWEVVEATCACLAAQEEAEKE-----HCSKC-----	269
FT	DISULFID	549	BY SIMILARITY.	566	DB	492	LPAGQIVNCDCNAAGTQGNACRRDPLRGVCVKPNRGAHCELCAFGFHGSPCHPCQCSS	551
FT	DISULFID	568	BY SIMILARITY.	577	QY	270	--LAQOMILEEFGRCCL	283
FT	DISULFID	580	BY SIMILARITY.	590	DB	552	PGVANSCLDPESGQCM	567
FT	DISULFID	593	BY SIMILARITY.	605	RESULT 5			
FT	DISULFID	595	BY SIMILARITY.	611	LMB1_MOUSE			
FT	DISULFID	613	BY SIMILARITY.	622	ID	LMB1_MOUSE	STANDARD;	PRT; 1786 AA.
FT	DISULFID	625	BY SIMILARITY.	635	AC	P02469;		
FT	DISULFID	638	BY SIMILARITY.	650	DT	21-JUL-1986	(Rel. 01, Created)	
FT	DISULFID	640	BY SIMILARITY.	656	DT	01-JUL-1989	(Rel. 11, Last sequence update)	
FT	DISULFID	658	BY SIMILARITY.	667	DT	15-JUN-2002	(Rel. 41, Last annotation update)	
FT	DISULFID	670	BY SIMILARITY.	680	DE	Laminin beta-1 chain precursor	(Laminin B1 chain).	
FT	DISULFID	683	BY SIMILARITY.	695	OS	LAMB1-1 OR LAMB-1.		
FT	DISULFID	685	BY SIMILARITY.	702	OS	Mus musculus (Mouse).		
FT	DISULFID	704	BY SIMILARITY.	713	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
FT	DISULFID	716	BY SIMILARITY.	726	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
FT	DISULFID	718	BY SIMILARITY.	728	OC	NCBI_TaxID=10090;		
FT	DISULFID	720	BY SIMILARITY.	730	RN	[1]		
FT	DISULFID	722	BY SIMILARITY.	732	RP	SEQUENCE FROM N.A.		
FT	DISULFID	724	BY SIMILARITY.	734	RX	MEDLINE=87147212; PubMed=3493487;		
FT	DISULFID	726	BY SIMILARITY.	736	RA	Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;		
FT	DISULFID	728	BY SIMILARITY.	738	RT	"Sequence of the cDNA encoding the laminin B1 chain reveals a		
FT	DISULFID	730	BY SIMILARITY.	740	RT	multidomain protein containing cysteine-rich repeats.";		
FT	DISULFID	732	BY SIMILARITY.	742	RL	Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).		
FT	DISULFID	734	BY SIMILARITY.	744	RN	[2]		
FT	DISULFID	736	BY SIMILARITY.	746	RP	SEQUENCE OF 1292-1786 FROM N.A.		
FT	DISULFID	738	BY SIMILARITY.	748	RX	MEDLINE=85051302; PubMed=6209134;		
FT	DISULFID	740	BY SIMILARITY.	750	RA	Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;		
FT	DISULFID	742	BY SIMILARITY.	752	RT	"Sequencing of laminin B chain cDNAs reveals C-terminal regions of		
FT	DISULFID	744	BY SIMILARITY.	754	RT	coiled-coil alpha-helix.";		
FT	DISULFID	746	BY SIMILARITY.	756	RL	EMBO J. 3:2355-2362(1984).		
FT	DISULFID	748	BY SIMILARITY.	758	RN	[3]		
FT	DISULFID	750	BY SIMILARITY.	760	RP	SEQUENCE OF 165-172; 539-547 AND 712-719.		
FT	DISULFID	752	BY SIMILARITY.	762	RC	STRAIN=BALEB/C; TISSUE=Endothelial cells;		
FT	DISULFID	754	BY SIMILARITY.	764	RX	MEDLINE=97363207; PubMed=9219532;		
FT	DISULFID	756	BY SIMILARITY.	766	RA	Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.;		
FT	DISULFID	758	BY SIMILARITY.	768	RT	Sorokin L.M.;		
FT	DISULFID	760	BY SIMILARITY.	770	RT	"Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of		
FT	DISULFID	762	BY SIMILARITY.	772	RL	endothelium.";		
FT	DISULFID	764	BY SIMILARITY.	774	RL	Eur. J. Biochem. 246:727-735(1997).		
FT	DISULFID	766	BY SIMILARITY.	776	CC	-!- FUNCTION: Binding to cells via a high affinity receptor, laminin		
FT	DISULFID	768	BY SIMILARITY.	778	CC	is thought to mediate the attachment, migration, and organization		
FT	DISULFID	770	BY SIMILARITY.	780	CC	of cells into tissues during embryonic development by interacting		
FT	DISULFID	772	BY SIMILARITY.	782	CC	with other extracellular matrix components.		
FT	DISULFID	774	BY SIMILARITY.	784	CC	-!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three		
FT	DISULFID	776	BY SIMILARITY.	786	CC	different polypeptide chains (alpha, beta, gamma), which are bound		
FT	DISULFID	778	BY SIMILARITY.	788	CC	to each other by disulfide bonds into a cross-shaped molecule		
FT	DISULFID	780	BY SIMILARITY.	790	CC	comprising one long and three short arms with globules at each		
FT	DISULFID	782	BY SIMILARITY.	792	CC	end.		
FT	DISULFID	784	BY SIMILARITY.	794	CC	THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-		
FT	DISULFID	786	BY SIMILARITY.	796	CC	2 (MEROSIN), AND LAMININ-6 (K-LAMININ).		
FT	DISULFID	788	BY SIMILARITY.	798	CC	-!- SUBCELLULAR LOCATION: Extracellular.		
FT	DISULFID	790	BY SIMILARITY.	800	CC	-!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR		
FT	DISULFID	792	BY SIMILARITY.	802	CC	COMPONENT).		
FT	DISULFID	794	BY SIMILARITY.	804	CC	-!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).		

Query Match 6.5%; Score 106.5; DB 1; Length 3718;
Best Local Similarity 21.9%; Pred. No. 0.67;
Matches 56; Conservative 26; Mismatches 87; Indels 87; Gaps 15;

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/PI50,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
 DE subunit).
 GN ITGB2 OR CD18.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee J.K., Schook L.B., Rutherford M.S.;
 RT "Molecular cloning and characterization of the porcine CD18 leukocyte adhesion molecule."
 RL xenotransplantation 3:222-230(1996).
 CC -!- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2, ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2 RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2 ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U13941; AAB16868.1; .
 DR HSP; P05106; IJV2.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002035; VWF A.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR ProDom; PD001811; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 769
 FT DOMAIN 23 700
 FT TRANSMEM 701 723
 FT DOMAIN 724 769
 FT DOMAIN 124 363
 FT DOMAIN 449 617
 FT REPEAT 449 496
 FT REPEAT 497 540
 FT REPEAT 541 581
 FT REPEAT 582 617
 FT REPEAT 618 647
 FT DISULFID 25 447
 FT DISULFID 33 43
 FT DISULFID 36 73
 FT DISULFID 46 62
 FT DISULFID 191 198
 FT DISULFID 246 286

FT DISULFID 386 400 BY SIMILARITY.
 FT DISULFID 420 562 BY SIMILARITY.
 FT DISULFID 445 449 BY SIMILARITY.
 FT DISULFID 459 470 BY SIMILARITY.
 FT DISULFID 467 506 BY SIMILARITY.
 FT DISULFID 472 481 BY SIMILARITY.
 FT DISULFID 483 497 BY SIMILARITY.
 FT DISULFID 512 517 BY SIMILARITY.
 FT DISULFID 514 549 BY SIMILARITY.
 FT DISULFID 519 534 BY SIMILARITY.
 FT DISULFID 536 541 BY SIMILARITY.
 FT DISULFID 557 562 BY SIMILARITY.
 FT DISULFID 559 590 BY SIMILARITY.
 FT DISULFID 564 573 BY SIMILARITY.
 FT DISULFID 575 582 BY SIMILARITY.
 FT DISULFID 596 601 BY SIMILARITY.
 FT DISULFID 598 643 BY SIMILARITY.
 FT DISULFID 603 612 BY SIMILARITY.
 FT DISULFID 615 618 BY SIMILARITY.
 FT DISULFID 622 631 BY SIMILARITY.
 FT DISULFID 628 695 BY SIMILARITY.
 FT DISULFID 647 670 BY SIMILARITY.
 FT SITE 397 399 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 50 50 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 642 642 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 769 AA; 84789 MW; FDD606CEEE850449 CRC64;
 Query Match 6.3%; Score 103; DB 1; Length 769;
 Best Local Similarity 21.5%; Pred. No. 0.25;
 Matches 59; Conservative 28; Mismatches 88; Indels 100; Gaps 16;
 QY 14 CIDNSTRRELKALHLVPOQNNYLDSDVPKPMALVGRFLPASTKLNLIITQLEGALP 73
 DB 449 CGDSKERTLCG-----NKGSMGCVCRDAGYIGKHCECOTQ-GRSQELEGSCR 498
 QY 74 SVVNGSAFPGSTLPGPKITLAGYCDCFAS-----GDFCNCCNCCNCCNLLHHDE 125
 DB 499 K-DNSSIICSGL-----GDCICGQCVCHTSDVPNKKYGFQFC-ECDNNMC-----E 542
 QY 126 RFAIKACLG-----RNPEAFQPKIGKGLGNVQHNKGCNRRS--GCLK-- 170
 DB 543 REDG-QVCGEGKRLGFCSTCRQSGFE-----GSACQCLKSTQGLNLQ 586
 QY 171 -----NYCEC--YE-----AQIMCSSICKICGCKNYEESPERKTLMSM 206
 DB 587 GVECSGRGRRCNVQCDFGYQPLCTDCPCQVPCARYAKCAEGLKFDTPGFAKNCSAE 646
 QY 207 PNYMTGLEGSHYLPPTKFGSLPRFSDRRPSSC 241
 DB 647 C-----GTTKULPSRMGS--RKCNERDSEGC 670
 RESULT 7
 LMB1_DROME
 ID LMB1_DROME STANDARD; PRT; 1790 AA.
 AC P11046; Q26328; Q9XZT4; Q9VLM6;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin beta-1 chain precursor (Laminin B1 chain).
 GN LAMB1 OR CG7123.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=94000382; PubMed=8397815;

RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
 RT "Analysis of the Drosophila gene for the laminin B1 chain";
 RL DNA Cell Biol. 12:573-587(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88210471; PubMed=3365769;
 RA Montell D.J., Goodman C.S.;
 RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
 RL reveals domains of homology with mouse";
 RN Cell 53:463-473(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.F., Benos P.V., Bertram M.R., Bouck J., Brokstein P., Brottier P.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RL "The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195(2000).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M95811; AAD19752.1; -
 DR EMBL; M19525; AAA28663.1; -
 DR EMBL; AE003618; AAF52563.1; -
 DR PIR; A28783; WMFEB1.
 DR HSP; P02488; ILE.
 DR FlyBase; FBgn0002527; LanB1.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00053; laminin_EGF; 13.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PRO0011; EGF_LAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR SMART; SM00180; EGF_Lam; 11.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 10.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.
 FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 290 356 LAMININ EGF-LIKE 1.
 FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
 FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
 FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 562 789 LAMININ DOMAIN IV.
 FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
 FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
 FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
 FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
 FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.
 FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
 FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
 FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
 FT DOMAIN 1191 1407 DOMAIN II.
 FT DOMAIN 1408 1434 DOMAIN ALPHA.
 FT DOMAIN 1435 1790 DOMAIN I.
 FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
 FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).
 FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
 FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
 FT DISULFID 290 299 BY SIMILARITY.
 FT DISULFID 292 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 334 356 BY SIMILARITY.
 FT DISULFID 357 366 BY SIMILARITY.
 FT DISULFID 359 384 BY SIMILARITY.
 FT DISULFID 387 396 BY SIMILARITY.
 FT DISULFID 399 417 BY SIMILARITY.
 FT DISULFID 420 433 BY SIMILARITY.
 FT DISULFID 422 448 BY SIMILARITY.
 FT DISULFID 450 459 BY SIMILARITY.
 FT DISULFID 462 477 BY SIMILARITY.
 FT DISULFID 480 493 BY SIMILARITY.
 FT DISULFID 482 500 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 514 528 BY SIMILARITY.
 FT DISULFID 791 803 BY SIMILARITY.
 FT DISULFID 793 810 BY SIMILARITY.
 FT DISULFID 812 821 BY SIMILARITY.
 FT DISULFID 824 836 BY SIMILARITY.
 FT DISULFID 839 851 BY SIMILARITY.
 FT DISULFID 841 858 BY SIMILARITY.

```
FT DISULFID 860 869 BY SIMILARITY.
FT DISULFID 872 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 887 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 995 1014 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1066 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1788 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.28; Score 102; DB 1; Length 1790;
Best Local Similarity 21.48; Pred. No. 0.75;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

Qy 97 GYCD-----CF--ASGDFCNNC-----NCNCC-----CNLLHHDIERFKA 129
Db 1005 GNCBQTCGLKCLYQTGDHCELCCKDGFGLAQNCQCECFGLGNTNTIAHCDRTG 1064
Qy 130 IKACLRNPEAFQPKIGKGLGNVKNPHNK-----GCNRRSGCLKNYCEYEAQIM 181
Db 1065 QCPCL-----PNVQGVRCDOCAENHWKIASGEGCSCNCDPIGALHEQCNSYTGQCQ 1116
Qy 182 CS---SICKCIGCK-NYEESPERKTLMSMPNMYMTGGLEGSHYLPPTKFS--GLPRFSHD 235
Db 1117 CKPGEGGRACQOAHYGNFNPK-----CQPCEDQFGAADFQCD 1157
Qy 236 RPPSCISWEVVEVTCACLLAQGEAEKCSKLAQEMILEEFGRCLS 284
Db 1158 RETGNCVCHGIGGYKCNACARGYQPPHCSPC-----GECFN 1196

RESULT 8
TFH2_MOUSE
ID TFH2_MOUSE STANDARD; PRT; 396 AA.
AC Q9JIB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE TFIIH basal transcription factor complex p44 subunit (Basic
DE transcription factor 2.44 kDa subunit) (BTF2-p44) (General
DE transcription factor IIF polypeptide 2).
GN GTF2H2 OR BTF2P44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20414747; PubMed=10958627;
RA Endrizzi M.G., Hadinoto V., Gowney J.D., Miller W., Dietrich W.F.;
RT "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
CC -1- FUNCTION: Component of the core-TFIIH basal transcription factor
CC involved in nucleotide excision repair (NER) of DNA and, when
CC complexed to CAK, in RNA transcription by RNA polymerase II. The
CC N-terminus interacts with and regulates XPD whereas an intact C-
CC terminus is required for a successful escape of RNAP II from the
CC promoter (By similarity).
CC -1- SUBUNIT: One of the six subunits forming the core-TFIIH basal
CC transcription factor. Interacts with XPB, XPD, p82 and p34 (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 C4-TYPE ZINC FINGER.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242432; AAR82753.1; .
DR MGI; MGI:1345669; Gtf2h2.
DR InterPro; IPR004595; Ssl1.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00327; VWA; 1.
DR TIGRFAMs; TIGR00622; ssl1; 1.
DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE; PS00089; ZF_RING_2; FALSE_NEG.
KW Transcription regulation; DNA repair; Nuclear protein; Zinc-finger.
FT ZN_FING 292 309 C4-TYPE.
FT ZN_FING 346 383 RING-TYPE.
FT DOMAIN 242 245 POLY-SER.
SQ SEQUENCE 396 AA; 44687 MW; 1BC57AFD464AA83F CRC64;

Query Match 6.28; Score 101.5; DB 1; Length 396;
Best Local Similarity 21.88; Pred. No. 0.16;
Matches 48; Conservative 27; Mismatches 72; Indels 73; Gaps 11;

Qy 2 VICQLKGTQMLCIDNSTRRELKALHLVPOYQDQNNYLQSDVPKPMALVGRFLPASTKL 61
Db 210 VLAREGTGYHVLDETHYKELLAAHHVSPPPASSSS-----ECSLRMGFPQHTIA 260
Qy 62 NLITQOLEGALPSVVVNGSAFPGSGTLPGPKITLAGYCDPASFAGFCNCCNCC----- 116
Db 261 SLSDQD---AKPS-FSMAHLDNNSTEPG---LTLGGY-----FCPCRAKYCELPVE 305
Qy 117 -----CNLLHH-----DIERFKAICACLRNPEAFQPKIGKGLGN 152
Db 306 CKICGLTLVSAPHILARSYHHLFPLDAFOEISLEEKYGRFCYGCQ-----GE 352
Qy 153 VKPQHNGCN-CRRSGCLKNYCEYEAQIMCSICKICGC 191
Db 353 LKDQHYVYCTVCQNVFCVD--CDVF-----VHDSLHCCFCGC 386

RESULT 9
BAR3_CHITE
ID BAR3_CHITE STANDARD; PRT; 1700 AA.
AC Q03376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiani ring protein 3 precursor.
GN BR3.
```

OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90172404; PubMed=1689777;
RA Paulson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RT repetitive structure split by many introns.";
RL J. Mol. Biol. 211:331-349(1990).
CC -!- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: SALIVARY GLAND.
CC -!- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52263; CAA36506.1; -
CC PIR; S08167; S08167.
CC HSSP; P15358; 1SK2.
CC InterPro: IPR004153; CXCXC-repeat.
CC Pfam; PF03128; CXCXC; 71.
KW Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
FT SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 1700;
Best Local Similarity 22.6%; Pred. No. 0.86; Indels 68; Gaps 15;
Matches 51; Conservative 20; Mismatches 87;

QY 99 CDFASGD-----FC-NNCNC-----NNCCNHLHDIERFAIKACILGRNPEAF 141
DB 198 CSCECKGDKGCKGSKWCKNCKICPTAEPAGGCSAPLAWDD--KCSACAPAKWEKK 255
QY 142 QPKIGKQLGNVPOHNGKNCRRSGCLN-----YCEYE-----AQIMCSSICK 187
DB 256 EKVESGKIWN---PNTCEGCAQLNCPDNKANKETCCQCEKVKKCGGVFCCKDCS 312
QY 188 CIGCKNYEESPERTLMSMPNYMTGLESHVLP---PTKFSGLPR----- 231
DB 313 CV-CPGGKD---KTCTAPQVY---DGVACSCSPVNMQRPADGCPRPQWDKEKRCCEC 365
QY 232 -FSHRRPSSCSVEVVEATCAC-----LLAQGEAEKHCES-KCL 270
DB 366 PVKHDCNGKV-WDETICQICPRDAPVCTACKERGECSECKCI 409

RESULT 10
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (SC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (SC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M62321; AAA45676.1; -
CC PIR; A39166; GNWVC3.
CC HSSP; P27958; 1A1V.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR001650; Helicase_C.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00998; HCV_RdRP; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC ProDom; PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 192 383

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 3011;
Best Local Similarity 24.8%; Pred. No. 1.6;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAFF-SGSLPGPKITLAGYDCDFASGDFCNCCNCCNNL-----HHDIERFKAI 130
DB 1424 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNCVTQTVDFSLDPTTFITITLP 1478
QY 131 KACLGRNPEAFQPKIGKGLG---NVKPOHNGKNCRRSGCLKN--YCECYEAQIMCSSI 185
DB 1479 QDAVSRTQR--RGRTGRGPGIYRFVAPCE-----RPSGMFDSVLCICYDA----- 1523
QY 186 KCIKCKNYEESPERKTLMSMNYAQTGGL-----EG-----SHYLPPTKF 226
DB 1524 ----GCAMYTELTP-AETTIVRLRAYMNTPGLPVCDHLEFWEGVFTGLTHIDAHFLSQTKQ 1578
QY 227 SG--LPRFSDRRPSSCISWEVEATCACLQAQGEAEKEHCKSL 270
DB 1579 SGENLP-----YLVAYQATVCARAQAPPSPWDQWKKCL 1611

RESULT 11

EZ_DROME
ID EZ_DROME STANDARD; PRT; 760 AA.
AC P42124;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enhancer of zeste protein.
E(2).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019309; PubMed=8413234;
RA Jones R.S., Gelbart W.M.;

RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax.";
RL Mol. Cell. Biol. 13:6357-6366(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97164720; PubMed=9012527;
RA Carrington E.A., Jones R.S.;
RT "The Drosophila Enhancer of zeste gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution.";
RL Development 122:4073-4083(1996).
CC -!- FUNCTION: NEGATIVELY REGULATES THE SEGMENT IDENTITY GENES OF THE ANTENNAPEDIA (ANT-C) AND BITHORAX (BX-C) GENE COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL: U00180; AAC46462.1; .
CC FlyBase: FBgn0000629; E(z).
CC InterPro: IPR001005; Myb_DNA_binding.
CC InterPro: IPR001214; SET.
CC Pfam: PF00856; SET; 1.
CC SMART: SM00395; SANT; 1.
CC SMART: SM00317; SET; 1.
CC PROSITE: PS0280; SET; 1.
CC Transcription regulation; Nuclear protein; DNA-binding; Repressor;
KW Developmental protein.
KW DOMAIN 505 510 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 538 619 CYS-RICH.
FT DOMAIN 632 743 SET.
SQ SEQUENCE 760 AA; 86965 MW; 4A28A8EAD7968AC4 CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 760;
Best Local Similarity 20.3%; Pred. No. 0.4;
Matches 36; Conservative 23; Mismatches 63; Indels 55; Gaps 8;

QY 99 CDCFAGDGFCCNN-CNCCNCCNNLHHDIERFKAIKACLRNPEAFQPKIGKGLGNVQKH 157
DB 549 CSCIQTQNECFECNCCSDCQN-----RP----- 572
QY 158 NKGCMCRRSGCLKNYCECYEAQIMCS-SICKICGCKNYEESPERKTLMSMNYMOTGGLE 216
DB 573 -PGCRC-KAQCNTKQCPYLAVRECDPDLQACGADQF-----KLTFKITCKNVQVQRLH 625
QY 217 GSHYLPPTKFSGLPFRSHDRPSSCISWEVEATCACLQAQ-----GGEAEKEHCKS 267
DB 626 KHLMLAPSDIAGWIGIFLKEGAQKN-----EFISEYCGEIIISQDEADRRRKVYDKYMC 678

RESULT 12
LMA5_HUMAN
ID LMA5_HUMAN STANDARD; PRT; 3695 AA.
AC O15230; Q9H1PI; Q8WZA7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5 OR KIAA0533 OR KIAA1907.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McTay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RN SEQUENCE OF 197-1934 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=21456161; PubMed=11572484;
RX Nagase T., Kikuno R., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
RN [3]
RN SEQUENCE OF 2051-3695 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=98290545; PubMed=9628581;
RX Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [4]
RN SEQUENCE OF 2743-3695 FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=97415425; PubMed=9271224;
RA Durkin M.E., Loechel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,
RA Wewer U.M.;
RT "Tissue-specific expression of the human laminin alpha5-chain, and
RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
RL FEBS Lett. 411:296-300(1997).
RN [5]
RN EXPRESSION IN RETINA.
RP MEDLINE=20422761; PubMed=10964957;
RA Libby R.T., Champiaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;
RT "Laminin expression in adult and developing retinae: evidence of two
RT novel CNS laminins.";
RL J. Neurosci. 20:6517-6528(2000).
CC [-] FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC [-] SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROTRIMER COMPOSED OF THREE
CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY
CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC [-] SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).

CC [-] TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN
CC BRAIN AND LIVER.
CC [-] DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROTRIMER.
CC [-] SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC [-] SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC [-] SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC [-] SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL354836; CAC22309.1; ALT_SEQ.
CC EMBL; AL354836; CAC22310.1;
CC EMBL; AB067494; BAB67800.1;
CC EMBL; AB011105; BAA25459.1;
CC EMBL; Z95636; CAB09137.1;
CC HSP; P02468; IKLO.
CC Genew; HGNC:6485; LAMA5.
CC MIM; 601033;
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00053; laminin_EGF; 2.
CC Pfam; PF00054; laminin_G; 2.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC PROSITE; PS00022; EGF_1; 19.
CC PROSITE; PS01186; EGF_2; 3.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC PROSITE; PS00025; LAM_G_DOMAIN; 5.
CC Glycoproteins; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL
CC 1 35 POTENTIAL.
CC FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
CC FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
CC FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
CC FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
CC FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
CC FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
CC FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
CC FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
CC FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
CC FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
CC FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
CC FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC FT DOMAIN 1831 1863 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC FT DOMAIN 1864 1912 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1913 1968 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1969 2022 LAMININ EGF-LIKE 19.
CC FT DOMAIN 2023 2069 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2070 2116 LAMININ EGF-LIKE 21.
CC FT DOMAIN 2117 2166 LAMININ EGF-LIKE 22.
CC FT DOMAIN 2167 2735 DOMAIN II AND I.
CC FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
CC FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
CC FT DOMAIN 3124 3292 LAMININ G-LIKE 3.


```

FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
FT DOMAIN 2201 2221 COILED COIL (POTENTIAL).
FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1838 1840 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 302 322 BY SIMILARITY.
FT DISULFID 324 333 BY SIMILARITY.
FT DISULFID 336 356 BY SIMILARITY.
FT DISULFID 359 368 BY SIMILARITY.
FT DISULFID 361 393 BY SIMILARITY.
FT DISULFID 396 405 BY SIMILARITY.
FT DISULFID 408 426 BY SIMILARITY.
FT DISULFID 429 440 BY SIMILARITY.
FT DISULFID 431 447 BY SIMILARITY.
FT DISULFID 449 458 BY SIMILARITY.
FT DISULFID 461 471 BY SIMILARITY.
FT DISULFID 494 506 BY SIMILARITY.
FT DISULFID 496 515 BY SIMILARITY.
FT DISULFID 517 526 BY SIMILARITY.
FT DISULFID 529 538 BY SIMILARITY.
FT DISULFID 541 553 BY SIMILARITY.
FT DISULFID 543 560 BY SIMILARITY.
FT DISULFID 562 571 BY SIMILARITY.
FT DISULFID 574 584 BY SIMILARITY.
FT DISULFID 587 599 BY SIMILARITY.
FT DISULFID 589 605 BY SIMILARITY.
FT DISULFID 607 616 BY SIMILARITY.
FT DISULFID 619 629 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT DISULFID 634 650 BY SIMILARITY.
FT DISULFID 652 661 BY SIMILARITY.
FT DISULFID 664 674 BY SIMILARITY.
FT DISULFID 677 689 BY SIMILARITY.
FT DISULFID 679 696 BY SIMILARITY.
FT DISULFID 698 707 BY SIMILARITY.
FT DISULFID 710 720 BY SIMILARITY.

Query Match 5.1%; Score 100; DB 1; Length 3695;
Best Local Similarity 21.8%; Pred. No. 2.4;
Matches 57; Conservative 23; Mismatches 100; Indels 82; Gaps 14;

QY 93 ITLAGVDCFCASGDFC-----NVC--NCNCCNLLHHDIEREKAIRA- 132
DB 294 ISIGRCVCHGADACDAKDPDPFLQCTCOHNTCGGTCDCRCPCFNQ--QFWKPATAN 351
QY 133 -----CLGR-----NPE-----AFQPKIGKQGLGNV---KQHNKNGCNCRRSGCL 169
DB 352 SANECQSCNCHYATDCYVDPEVDRRRASQSLDGYQGGVCIDCQHHHTGVNCR--CL 409
QY 170 KNYCEYENQINCSICKICGKN-----YESPERKTLMSPNY-----MOTGGLEG 217
DB 410 PGFYRSPNHLPSHVCRNCESDFTDGTCDLTGR--CYCRPNFSGRCVCAEGFTG 467
QY 218 SHVLPTRFSGGLPRFSDRRPSSCTSWVEVATCACLLAQGEAEKEKCKLAQEMILE 277
DB 468 FPCSTPT-----PSSNDTREQVLFPAGQIVNCDCSAAGTQGNACRKD-----P 510
QY 278 EFGRCLS-----QILHTEFKSKG 295
DB 511 RVGRCLKPNFQGTCELCAFG 532

RESULT 13
ID ITB6_MOUSE STANDARD; PRT: 787 AA.
AC Q920T9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin beta-6 precursor.
```

```

GN ITGB6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20547423; PubMed=11095652;
RA Arend L.J., Smart A.M., Briggs J.P.;
RT "Mouse beta(6) integrin sequence, pattern of expression, and role in
RT kidney development.";
RL J. Am. Soc. Nephrol. 11:2297-2305(2000).
CC !- FUNCTION: INTEGRIN ALPHA-V/BETA-6 IS A RECEPTOR FOR FIBRONECTIN
CC AND CYTOTACTIN. IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGANDS
CC (BY SIMILARITY).
CC !- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-6
CC ASSOCIATES WITH ALPHA-V (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC !- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF115376; AAD17212.1;
CC HSSP: P05106; 1JV2.
CC MGD: MGI:96615; Itgb6
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002369; Integrin_B.
CC InterPro: IPR001169; Integrin_beta_C.
CC InterPro: IPR003659; Plexin-like.
CC InterPro: IPR002035; VWF_A.
CC Pfam: PF00362; Integrin_B; 1.
CC PRINTS: PR01186; INTEGRINB.
CC ProDom: PM001811; Integrin_B; 1.
CC SMART: SM00001; EGF_like; 1.
CC SMART: SM00187; INB; 1.
CC SMART: SM00423; PSI; 1.
CC SMART: SM00327; VWA; 1.
CC PROSITE: PS00243; INTEGRIN_BETA; 2.
CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
CC PROSITE: PS01186; EGF_2; UNKNOWN_1.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 787 INTEGRIN BETA-6.
FT DOMAIN 22 706 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 707 729 POTENTIAL.
FT DOMAIN 730 787 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 131 371 VWFA-LIKE.
FT DOMAIN 456 619 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 456 501 I.
FT REPEAT 502 543 II.
FT REPEAT 544 582 III.
FT REPEAT 583 619 IV.
FT DISULFID 23 454 BY SIMILARITY.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 34 70 BY SIMILARITY.
FT DISULFID 44 59 BY SIMILARITY.
FT DISULFID 197 204 BY SIMILARITY.
FT DISULFID 252 293 BY SIMILARITY.
FT DISULFID 394 406 BY SIMILARITY.
FT DISULFID 426 669 BY SIMILARITY.
FT DISULFID 452 456 BY SIMILARITY.
FT DISULFID 467 479 BY SIMILARITY.
FT DISULFID 476 511 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
```


DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
FT DOMAIN 23 282 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1376 1415 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.
FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
FT DOMAIN 1570 2140 DOMAIN II AND I.
FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
FT DOMAIN 2325 2517 LAMININ G-LIKE 2.
FT DOMAIN 2522 2706 LAMININ G-LIKE 3.
FT DOMAIN 2709 2930 LAMININ G-LIKE 4.
FT DOMAIN 2929 3106 LAMININ G-LIKE 5.
FT DOMAIN 1662 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1823 2146 COILED COIL (POTENTIAL).
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 285 303 BY SIMILARITY.
FT DISULFID 305 314 BY SIMILARITY.
FT DISULFID 317 337 BY SIMILARITY.
FT DISULFID 340 349 BY SIMILARITY.
FT DISULFID 342 374 BY SIMILARITY.
FT DISULFID 377 386 BY SIMILARITY.
FT DISULFID 389 407 BY SIMILARITY.
FT DISULFID 410 422 BY SIMILARITY.
FT DISULFID 412 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 452 462 BY SIMILARITY.
FT DISULFID 465 478 BY SIMILARITY.
FT DISULFID 467 482 BY SIMILARITY.
FT DISULFID 484 493 BY SIMILARITY.
FT DISULFID 496 511 BY SIMILARITY.
FT DISULFID 513 511 BY SIMILARITY.
FT DISULFID 753 762 BY SIMILARITY.
FT DISULFID 755 769 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 784 800 BY SIMILARITY.
FT DISULFID 803 818 BY SIMILARITY.
FT DISULFID 805 828 BY SIMILARITY.
FT DISULFID 831 840 BY SIMILARITY.
FT DISULFID 843 858 BY SIMILARITY.
FT DISULFID 861 875 BY SIMILARITY.
FT DISULFID 883 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 897 911 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 916 933 BY SIMILARITY.
FT DISULFID 935 944 BY SIMILARITY.
FT DISULFID 947 960 BY SIMILARITY.
FT DISULFID 963 975 BY SIMILARITY.
FT DISULFID 965 981 BY SIMILARITY.
FT DISULFID 983 992 BY SIMILARITY.
FT DISULFID 995 1007 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.

FT DISULFID 1012 1026 BY SIMILARITY.
FT DISULFID 1028 1037 BY SIMILARITY.
FT DISULFID 1040 1053 BY SIMILARITY.
FT DISULFID 1056 1068 BY SIMILARITY.
FT DISULFID 1077 1086 BY SIMILARITY.
FT DISULFID 1089 1099 BY SIMILARITY.
FT DISULFID 1416 1425 BY SIMILARITY.
FT DISULFID 1418 1432 BY SIMILARITY.
FT DISULFID 1435 1444 BY SIMILARITY.
FT DISULFID 1447 1462 BY SIMILARITY.
FT DISULFID 1465 1480 BY SIMILARITY.
FT DISULFID 1467 1490 BY SIMILARITY.
FT DISULFID 1493 1502 BY SIMILARITY.
FT DISULFID 1505 1520 BY SIMILARITY.
FT DISULFID 1523 1535 BY SIMILARITY.
FT DISULFID 1525 1542 BY SIMILARITY.
FT DISULFID 1544 1553 BY SIMILARITY.
FT DISULFID 1556 1567 BY SIMILARITY.
FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
FT DISULFID 1574 1574 INTERCHAIN (PROBABLE).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
Query Match 6.0%; Score 98.5; DB 1; Length 3106;
Best Local Similarity 22.0%; Pred. No. 2.7;
Matches 70; Conservative 21; Mismatches 82; Indels 145; Gaps 19;
Qy 73 PSVNGSAPP-----SGSTL---PGPPKITLAGYDCDFAS---GDFFC 108
Db 851 PSVPGGSCQPCQCNLDYSPGSCDSLSGCLCKPG---TTRYCELCDAGVFGDAV 906
Qy 109 NNCNCCN-CH-----NLHHDIERFAIKRACLRGN-----PEAFQPKRIGKQLGNVKNP 155
Db 907 NTKNCQPCRCIDINGSFSDCHTRTGQCECRPNVQGRHCDCKPETFGLQGRGCL----- 961
Qy 156 QHNKGCNRRSGCLKNYCEYEAQIMCS---SICKICGCK-----NYEE----- 196
Db 962 ---PCNCSFGSKSFDCEA---SGQCWCQPGVAGKKCDCAHGYNFQEGGCIACDCSHLG 1016
Qy 197 ---SPERTKLSMPNMYMOTGLEGSHYLPPT----- 224
Db 1017 NNCDPRTQCICPPN---TTGEKCECLPNTWGHISIVTGCKVCNCSVTGSLASQCQNVNVTG 1073
Qy 225 -----KFSGLPRFSDHRRPSSCI---SWEV-----VEAT-----CACLLA 256
Db 1074 QCSCHPKFSGM-----KCSECSRGHWNYPICLTCLDCFLPCTDATTCDLETRKCSQDQ 1126
Qy 257 QGE-----EAEKEHCKSC 269
Db 1127 TQCCKVNVGVGHCDRC 1144
RESULT 15
ZAN_MOUSE
ID ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC O88799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;

RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
FT membrane protein containing multiple cell adhesion molecule-like
RL domains.";
RJ J. Biol. Chem. 273:3415-3421(1998).
[2]
RN SEQUENCE OF 4864-5376 FROM N.A.
RP TISSUE=Testis;
RC MEDLINE=97271566; PubMed=9126492;
RX Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RL zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -1- DOMAIN: THE VWF DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 25 VWF DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U97068; AAC26680.1; -
DR EMBL; U83190; AAC53125.1; -
DR MGD; MGI:108656; Zan.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILa_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00094; vwd; 4.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILa; 25.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FOLN; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS50060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 18 5310. EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337 POTENTIAL.
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 210 MAM 1.
FT DOMAIN 215 374 MAM 2.
FT DOMAIN 377 542 MAM 3.

FT	DOMAIN	547	1170	80 X HEPTAPEPTIDE REPEATS (APPROXIMATE) (MUCIN-LIKE DOMAIN).
FT	DOMAIN	1171	1280	VWFD 1 (PARTIAL).
FT	DOMAIN	1281	1669	VWFD 2.
FT	DOMAIN	1670	2056	VWFD 3.
FT	DOMAIN	2057	2459	VWFD 4.
FT	DOMAIN	2460	2579	VWFD 5 (PARTIAL).
FT	DOMAIN	2580	2699	VWFD 6 (PARTIAL).
FT	DOMAIN	2700	2819	VWFD 7 (PARTIAL).
FT	DOMAIN	2820	2939	VWFD 8 (PARTIAL).
FT	DOMAIN	2940	3059	VWFD 9 (PARTIAL).
FT	DOMAIN	3060	3179	VWFD 10 (PARTIAL).
FT	DOMAIN	3180	3299	VWFD 11 (PARTIAL).
FT	DOMAIN	3300	3416	VWFD 12 (PARTIAL).
FT	DOMAIN	3417	3536	VWFD 13 (PARTIAL).
FT	DOMAIN	3537	3656	VWFD 14 (PARTIAL).
FT	DOMAIN	3657	3776	VWFD 15 (PARTIAL).
FT	DOMAIN	3777	3892	VWFD 16 (PARTIAL).
FT	DOMAIN	3893	4928	VWFD 17 (PARTIAL).
FT	DOMAIN	4029	4148	VWFD 18 (PARTIAL).
FT	DOMAIN	4149	4263	VWFD 19 (PARTIAL).
FT	DOMAIN	4264	4283	VWFD 20 (PARTIAL).
FT	DOMAIN	4384	4503	VWFD 21 (PARTIAL).
FT	DOMAIN	4504	4623	VWFD 22 (PARTIAL).
FT	DOMAIN	4624	4743	VWFD 23 (PARTIAL).
FT	DOMAIN	4744	4863	VWFD 24 (PARTIAL).
FT	DOMAIN	4864	5261	VWFD 25.
FT	DOMAIN	5259	5295	EGF-LIKE.
FT	DISULFID	5263	5274	BY SIMILARITY.
FT	DISULFID	5268	5283	BY SIMILARITY.
FT	DISULFID	5285	5294	BY SIMILARITY.
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1216	1216	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1314	1314	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1814	1814	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	5376 AA;	579908 MW;	0B44DB77DF2A2620 CRC64;

Query Match 6.0%; Score 98.5; DB 1; Length 5376;

Best Local Similarity 23.2%; Pred. No. 4.8;

Matches 53; Conservative 19; Mismatches 73; Indels 83; Gaps 15;

QY 99 CDCFASGDFCNCCNC-----NCCNNLHDIERFAKAKL-GRNPEAFQPKIGKQL 150

DB 4232 CACVGAQCCHNETCTGTGTCQNSCKITVQCFAHSQYITCLPCLPSCFDP---EGLC 4288

Oy 151 GNVKPOH-----NKGNC-----RRSG-LKNYCEYEAQ-----IMCSSICK 187
| | : | | | | | : | | : | |
Db 4289 GGASPRAPSTCREGCVCEADYVLRREDKCVLRTOGCKDAQGLIPANKTWTLTRGCAOKCT 4348
| | : | | | | | : | | : | |
Oy 188 CIG-----CKNYE-----ESPERKTLMSMPNMYMOTGGLEGSHY---LPPTKFS 227
| | : | | | | | : | | : | |
Db 4349 CKGGNIHCWNFKPLGTECKDSVDGGSNCTKIALOCPAH-----SHHTYCLP-----S 4396
| | : | | | | | : | | : | |
Oy 228 GLPRFS---HDR-----RPSSCISWEVVEATCACLLAQGEAEKEHC 266
| | : | | | | | : | | : | |
Db 4397 CIPSCSNVNDRCSTSLQRPSTCIE-----GCLCHSGFVFSKDKC 4436
| | : | | | | | : | | : | |

Search completed: April 21, 2003, 11:49:02
Job time : 15.0606 secs

Qy 121 ERFKAIAKLDNRNPAFQPMKGRGLGAALRHSGKNCGRSGCLKNYCEYEAKIMCSS 180
 |||||
 Db 121 ERFKAIAKLDNRNPAFQPMKGRGLGAALRHSGKNCGRSGCLKNYCEYEAKIMCSS 180
 |||||
 Qy 181 IKCIIACKNYEESPRKMLMSTPHYMEPDGDESSHYLSPAKFSPPKLRKRNQAFSCISW 240
 |||||
 Db 181 IKCIIACKNYEESPRKMLMSTPHYMEPDGDESSHYLSPAKFSPPKLRKRNQAFSCISW 240
 |||||
 Qy 241 EWVATCACLAQGEAEQEHCSPLAEQMILEEFGRCLSOILHTEFKSKGLKIE 295
 |||||
 Db 241 EWVATCACLAQGEAEQEHCSPLAEQMILEEFGRCLSOILHTEFKSKGLKIE 295
 |||||

RESULT 2

MTL5_HUMAN STANDARD; PRT; 299 AA.
 AC Q9Y4I5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific
 metallothionein-like protein).
 DE MTL5.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUP-Testis;
 RX MEDLINE=99208669; PubMed=10191092;
 RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
 RT "A novel testis-specific metallothionein-like protein, tesmin, is an
 early marker of male germ cell differentiation.";
 RL Genomics 57:130-136(1999).
 CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U86074; AD24668.1; -;
 CC Genew; HGNC:7446; MTL5.
 CC MIM; 604374; -;
 CC InterPro; IPR005172; CXC.
 CC Pfam; PF03638; CXC; 1.
 CC Spermatogenesis.
 FT DOMAIN 99 191 CYS-RICH.
 SQ SEQUENCE 299 AA; 33110 MW; 8f679e140f969c11 CRC64;

Query Match 78.28; Score 1251; DB 1; Length 299;
 Best Local Similarity 76.38; Pred. No. 2.6e-101;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICQLKGGAGMLCNDGARELKHLLPQYDQSSFPQSELPKPMTLVGRLLPVPK 60
 |||||
 Db 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYLSQSDVPKPMALVGRPLPASTK 60
 |||||
 Qy 61 LNLITQVNGALPSAVNGAAPPSPALOGPPKPYLISGVCDFSGDFCNCS- - - - -CNL 116
 |||||
 Db 61 LNLITQLEGALPSVNGAAPPSPSTLPFPKPYLAGYCDCAFSGDFCNCCNCCN 120
 |||||
 Qy 117 RHELEKAIKACIDRNPAPQPMKGRGLGAALRHSGKNCGRSGCLKNYCEYEAKI 176
 |||||
 Db 121 HHDIERFAIKACILGRNPAFQPMKGRGLGAALRHSGKNCGRSGCLKNYCEYEAKI 180
 |||||
 Qy 177 MCSICKICKAKNYEESPERKMLMSTPHYMEPDGDESSHYLSPAKFSPPKLRKRNQAFS 236
 |||||

Db 181 MCSICKICKAKNYEESPERKMLMSTPHYMEPDGDESSHYLSPAKFSPPKLRKRNQAFS 240
 |||||
 Qy 237 CISWEVVEATCACLAQGEAEQEHCSPLAEQMILEEFGRCLSOILHTEFKSKGLKIE 295
 |||||
 Db 241 CISWEVVEATCACLAQGEAEQEHCSPLAEQMILEEFGRCLSOILHTEFKSKGLKME 299
 |||||
 RESULT 3
 RECK_HUMAN STANDARD; PRT; 971 AA.
 AC O95980; Q8WX37;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reversion-inducing cysteine-rich protein with Kazal motifs precursor
 DE (hRECK) (Suppressor of tumorigenicity 15) (ST15).
 GN RECK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
 RC TISSUP-Fibroblast;
 RX MEDLINE=99007295; PubMed=9789069;
 RA Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K.,
 RA Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y.,
 RA Ratzkin B.J., Arakawa T., Noda M.;
 RT "Regulation of matrix metalloproteinase-9 and inhibition of tumor
 invasion by the membrane-anchored glycoprotein RECK.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13221-13226(1998).
 RN [2]
 RP SEQUENCE OF 363-971 FROM N.A.
 RA Kimberley A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
 CC by suppressing MMP-9 secretion and by direct inhibition of its
 CC enzymatic activity. RECK down-regulation by oncogenic signals may
 CC facilitate tumor invasion and metastasis. Appears to also
 CC regulate MMP-2 and MT1-MMP, which are involved in cancer
 CC progression.
 CC -1- SUBUNIT: Interacts with MMP-9.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
 CC cells. It is undetectable in tumor-derived cell lines and
 CC oncogenically transformed cells.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D50406; BAA34060.1; -;
 CC EMBL; AL158830; CADI3384.1; -;
 CC Genew; HGNC:11345; RECK.
 CC MIM; 605227; -;
 CC HSSP; P80424; IAN1.
 CC InterPro; IPR002350; kazal.
 CC SMART; SM00280; KAZAL; 3.
 CC SMART; SM00011; VWC_def; 1.
 CC PROSITE; PS00282; KAZAL; 1.
 CC Signal; Glycoprotein; GPI-anchor; Serine protease inhibitor;
 CC Membrane; Anti-oncogene; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 942 REVERSION-INDUCING CYSTEINE-RICH PROTEIN
 FT WITH KAZAL MOTIFS.
 FT PROPEP 943 971 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 942 942 GPI-ANCHOR (POTENTIAL).


```

FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA: 68781 MW: 94E4C933AD943DB6 CRC64;

Query Match 6.5% Score 103.5; DB 1; Length 643;
Best Local Similarity 29.4%; Pred. No. 0.16;
Matches 37; Conservative 14; Mismatches 54; Indels 21; Gaps 6;

QY 83 SGPALQGPPTLS---GYC--DCSSGDFCNCSNLRHELEFRKAICACLDNRPEAF 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 SGP-LCVSPKFCGFSNNGCQDCFEFGGSPRCG---RPGFRLDLLVTCASRPGSS 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 138 OPKMGKGRGLAKLRHSGCKRSGCLKNYCEKVEAKIMSSICKTACKNYEESPERK 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 NPCTGGGMCHSVPLENTYCHCPRGYQLDS-----SQVHCVDIDEC-----EDSPCDQ 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 198 MLMSTP 203
   :|||
Db 354 ECINTP 359

RESULT 5
ID LM1 DROME STANDARD; PRT; 1790 AA.
AC P11046; Q26328; Q9X2T4; Q9VLW6;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LMNB1 OR LMNB1 OR CG7123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=94000382; PubMed=8397815;
RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
RT "Analysis of the Drosophila gene for the laminin B1 chain.";
RL DNA Cell Biol. 12:573-587(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88210471; PubMed=3365769;
RA Montell D.J., Goodman C.S.;
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson K.V., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zheng G., Zhou Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M95811; AAD19752.1; -
CC EMBL; M19525; AAA28663.1; -
CC EMBL; AE003618; AAP52563.1; -
CC PIR; A28783; MMFFB1.
CC HSSP; P02468; 1TLE.
CC Flybase; FBgn0002527; Lanbl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR002049; Laminin_EGF.
CC Pfam; PF00053; laminin_EGF; I3.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; LamNT; 1.
CC SMART; SM00180; EGF_Lam; 11.
CC SMART; SM00001; EGF_Like; 1.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 10.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 25 LAMININ BETA-1 CHAIN.
CC CHAIN 26 1790 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 43 289 LAMININ EGF-LIKE 1.
CC DOMAIN 290 356

```

FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
 FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
 FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 562 789 LAMININ DOMAIN IV.
 FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
 FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
 FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
 FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
 FT DOMAIN 1044 LAMININ EGF-LIKE 10.
 FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
 FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
 FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
 FT DOMAIN 1191 1407 DOMAIN II.
 FT DOMAIN 1408 1434 DOMAIN ALPHA.
 FT DOMAIN 1435 1700 DOMAIN I.
 FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
 FT DOMAIN 1435 1507 COILED COIL (POTENTIAL).
 FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
 FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
 FT DISULFID 290 299 BY SIMILARITY.
 FT DISULFID 292 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 334 354 BY SIMILARITY.
 FT DISULFID 357 366 BY SIMILARITY.
 FT DISULFID 359 384 BY SIMILARITY.
 FT DISULFID 387 396 BY SIMILARITY.
 FT DISULFID 399 417 BY SIMILARITY.
 FT DISULFID 420 433 BY SIMILARITY.
 FT DISULFID 422 448 BY SIMILARITY.
 FT DISULFID 450 459 BY SIMILARITY.
 FT DISULFID 462 477 BY SIMILARITY.
 FT DISULFID 480 493 BY SIMILARITY.
 FT DISULFID 482 500 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 514 528 BY SIMILARITY.
 FT DISULFID 791 803 BY SIMILARITY.
 FT DISULFID 793 810 BY SIMILARITY.
 FT DISULFID 812 821 BY SIMILARITY.
 FT DISULFID 824 836 BY SIMILARITY.
 FT DISULFID 839 851 BY SIMILARITY.
 FT DISULFID 841 858 BY SIMILARITY.
 FT DISULFID 860 869 BY SIMILARITY.
 FT DISULFID 872 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 887 901 BY SIMILARITY.
 FT DISULFID 904 913 BY SIMILARITY.
 FT DISULFID 916 932 BY SIMILARITY.
 FT DISULFID 935 951 BY SIMILARITY.
 FT DISULFID 937 962 BY SIMILARITY.
 FT DISULFID 964 973 BY SIMILARITY.
 FT DISULFID 976 990 BY SIMILARITY.
 FT DISULFID 993 1007 BY SIMILARITY.
 FT DISULFID 995 1014 BY SIMILARITY.
 FT DISULFID 1017 1026 BY SIMILARITY.
 FT DISULFID 1029 1042 BY SIMILARITY.
 FT DISULFID 1045 1059 BY SIMILARITY.
 FT DISULFID 1047 1056 BY SIMILARITY.
 FT DISULFID 1068 1077 BY SIMILARITY.
 FT DISULFID 1080 1093 BY SIMILARITY.
 FT DISULFID 1096 1108 BY SIMILARITY.
 FT DISULFID 1098 1115 BY SIMILARITY.
 FT DISULFID 1117 1126 BY SIMILARITY.
 FT DISULFID 1129 1141 BY SIMILARITY.
 FT DISULFID 1144 1156 BY SIMILARITY.
 FT DISULFID 1146 1163 BY SIMILARITY.
 FT DISULFID 1165 1174 BY SIMILARITY.
 FT DISULFID 1177 1188 BY SIMILARITY.
 FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
 FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
 FT DISULFID 1788 1798 INTERCHAIN (PROBABLE).
 FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
 Best Local Similarity 21.0%; Pred. No. 2;
 Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TLSGYCDYCFSSGDF-----CNCSGC-----NNLRHELEFRKFAKRACLDRNPEAFQPKM 141
 Db 1021 TTGDHCELCKDGFGLDALQNCQCCDFLGTNTTAHCDRFTGQCPCPCL-----PNV 1072
 QY 142 GKRLGAALKLRHSK-----GCNKRSCGLKNYCEYAKIMCSSICIKACK-----N 189
 Db 1073 QGVRCDCQCAENHWKIASGEGCESCNDPIGALHEQCNSYTGQCCKPFGGRACNQCAH 1132
 QY 190 YEESPERKMLMSTPHYMEPCDFESSHYLSPAKESGPPKLRKNQAFCISWEVVEATCAC 249
 Db 1133 YWGNPNEKQPCEDQFGAADFQC-----DRETGNVCVCHGEGGYKCN 1175
 QY 250 LLAQGEAEQEHCS 264
 Db 1176 ECARGYIGQPHCSP 1190

RESULT 6
 MSL2_DROME STANDARD; PRT; 773 AA.
 AC P50534;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Male-specific lethal-2 protein.
 GN MSL-2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95300219; PubMed=7781064;
 RA Kuroda M.I.;
 RT "Expression of msl-2 causes assembly of dosage compensation
 regulators on the X chromosomes and female lethality in Drosophila."
 RL Cell 81:867-877(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95317307; PubMed=7796814;
 RA Zhou S., Yang Y., Scott M.J., Pannuti A., Fehr K.C., Eisen A.,
 Koonin E.V., Fouts D.L., Wrightsman R., Manning J.E., Lucchesi J.C.;
 RT "Male-specific lethal 2, a dosage compensation gene of Drosophila,
 undergoes sex-specific regulation and encodes a protein with a RING
 finger and a metallothionein-like cysteine cluster."
 RL EMBO J. 14:2884-2895(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96017637; PubMed=7588059;
 RA Bashaw G.J., Baker B.S.;
 RT "The msl-2 dosage compensation gene of Drosophila encodes a putative
 DNA-binding protein whose expression is sex specifically regulated
 by Sex-lethal."
 RL Development 121:3245-3258(1995).
 CC -!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
 TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
 DOSAGE COMPENSATION). MSL-2 IS REQUIRED FOR TRANSLATION AND/OR
 STABILITY OF MSL-1 IN MALES. BOTH MSL-1 AND MSL-2 BIND TO HUNDREDS
 OF SITES ALONG THE MALE X CHROMOSOME, BUT NOT THE FEMALE X OR THE

```
CC AUTOSOMES IN EITHER SEX.
CC -1- SUBUNIT: MSL-2 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-1.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: THE MSL-2 PROTEIN IS ONLY PRODUCED IN MALES.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L42553; AAA75573.1; .
DR EMBL; Z48443; CA88358.1; .
DR EMBL; X89241; CA61529.1; .
DR Flybase; FBgn0005616; msl-2.
DR InterPro; IPR001841; Znf_Ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING.1; 1.
DR PROSITE; PS00089; ZF_RING.2; 1.
KW Zinc-finger; Nuclear protein; DNA-binding; Coiled coil.
FT ZN_FING 41 85
FT DOMAIN 283 519
FT CONFLICT 280 280 T -> N (IN REF. 2).
FT CONFLICT 370 373 MISSING (IN REF. 2).
FT CONFLICT 371 373 LKT -> VEE (IN REF. 3).
FT CONFLICT 380 380 P -> Q (IN REF. 2).
FT CONFLICT 403 403 T -> I (IN REF. 2).
FT CONFLICT 405 405 T -> A (IN REF. 2).
FT CONFLICT 423 423 H -> Q (IN REF. 3).
FT CONFLICT 429 429 Q -> H (IN REF. 2).
FT CONFLICT 431 431 A -> E (IN REF. 2 AND 3).
FT CONFLICT 433 433 S -> P (IN REF. 2 AND 3).
FT CONFLICT 439 439 E -> V (IN REF. 2).
FT CONFLICT 443 443 A -> P (IN REF. 2).
FT CONFLICT 471 471 T -> A (IN REF. 3).
FT CONFLICT 591 591 M -> L (IN REF. 3).
FT CONFLICT 642 642 E -> D (IN REF. 2).
SQ SEQUENCE 773 AA; 84841 MW; AE38BA76A4174225 CRC64;

Query Match 6.08; Score 96; DB 1; Length 773;
Best Local Similarity 25.7%; Pred. No. 0.89;
Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

Oy 121 ERFAIKAC----LDRNEAFQPMKGRGLGAALRHSGK-----CNCKRSG 163
Db 475 KRTRLKASQAQAKTEPVSEVKTQVSGK-CA--LRRIRGDKDEKVKPPKRCGSG 531
Oy 164 -----CLKNYCEYEAKIMCSSICKCIACKN-----YEE-----PER 196
Db 532 SSNTLTTCRNSRCPCYKSYNSCAG-CHCVCKNPKHKEDYVESDEDDLEDFEMPDKVPEP 590
Oy 197 KMLMSTPHYMPGDFESHYLSPAKFSPPKL 228
Db 591 MTQSEPVVAEPPROEENS--MAPPDSSAPISL 620

RESULT 7
LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC Q61001; Q9JH06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpl R., Sasaki T.;
RT Completion of the N-terminal sequence of the murine Laminin alpha 5
RT chain.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT Molecular cloning of a novel laminin chain, alpha 5, and widespread
RT expression in adult mouse tissues.;
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ293593; CAB99255.1; .
DR EMBL; U37501; AAC53430.1; .
DR HSSP; P02468; 1TLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.
DR Pfam; PF00055; laminin_Nterm; 1.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE; PS50025; LAM_G_DOMAIN; 5.
DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal lung;
 RA MEDLINE=93300971; PubMed=7781776;
 RX Iivanainen A., Sainio K., Sariola H., Tryggvason K.;
 RT "Primary structure and expression of a novel human laminin alpha 4
 chain.";
 RL FEBS Lett. 365:183-188(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97454279; PubMed=9310354;
 RA Richards A.J., Luccarini C., Pope F.M.;
 RT "The structural organisation of LAMA4, the gene encoding laminin
 alpha4.";
 RL Eur. J. Biochem. 248:15-23(1997).
 RN [3]
 RP SEQUENCE OF 236-1816 FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=95048381; PubMed=7959779;
 RA Richards A.J., Al-Imara L., Carter N.P., Lloyd J.C., Leversha M.A.,
 Pope F.M.;
 RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of
 a partial cDNA encoding a variant laminin A chain.";
 RL Genomics 22:237-239(1994).
 RN [4]
 RP SEQUENCE OF 66-1816 FROM N.A.
 RA Tubby B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,
 OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO
 EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,
 BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN
 FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS
 IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- SIMILARITY: CONTAINS 3.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -1- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S78569; AAB34635.1; -;
 DR EMBL; X91171; CAA62596.1; -;
 DR EMBL; Y14240; CAA74636.1; -;
 DR EMBL; X76939; CAA54258.1; -;
 DR EMBL; Z99289; CAB16553.1; -;
 DR HSSP; P02468; IKLO.
 DR Genew; HGNC:6484; LAMA4.
 DR MIM; 600133; -;
 DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00053; laminin_EGF; 6.
 DR Pfam; PF00054; laminin_G; 6.
 DR SMART; SM00180; EGF_Lam; 3.
 DR SMART; SM00282; LamG; 5.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.
 FT DOMAIN 82 131 LAMININ EGF-LIKE 1.
 FT DOMAIN 132 186 LAMININ EGF-LIKE 2.
 FT DOMAIN 187 240 LAMININ EGF-LIKE 3.
 FT DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 256 825 DOMAIN II AND I.
 FT DOMAIN 826 1028 LAMININ G-LIKE 1.
 FT DOMAIN 1040 1220 LAMININ G-LIKE 2.
 FT DOMAIN 1227 1395 LAMININ G-LIKE 3.
 FT DOMAIN 1462 1633 LAMININ G-LIKE 4.
 FT DOMAIN 1640 1813 LAMININ G-LIKE 5.
 FT DOMAIN 313 396 COILED COIL (POTENTIAL).
 FT DOMAIN 466 521 COILED COIL (POTENTIAL).
 FT DOMAIN 574 607 COILED COIL (POTENTIAL).
 FT DOMAIN 655 717 COILED COIL (POTENTIAL).
 FT DOMAIN 779 799 COILED COIL (POTENTIAL).
 FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 82 91 BY SIMILARITY.
 FT DISULFID 84 98 BY SIMILARITY.
 FT DISULFID 101 110 BY SIMILARITY.
 FT DISULFID 113 129 BY SIMILARITY.
 FT DISULFID 132 146 BY SIMILARITY.
 FT DISULFID 134 155 BY SIMILARITY.
 FT DISULFID 157 186 BY SIMILARITY.
 FT DISULFID 169 184 BY SIMILARITY.
 FT DISULFID 187 202 BY SIMILARITY.
 FT DISULFID 189 209 BY SIMILARITY.
 FT DISULFID 212 221 BY SIMILARITY.
 FT DISULFID 224 238 BY SIMILARITY.
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 269 269 INTERCHAIN (PROBABLE).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1086 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1281 1281 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1359 1359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1411 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 143 143 A -> P (IN REF. 1).
 FT CONFLICT 178 178 L -> F (IN REF. 1).
 FT CONFLICT 265 265 G -> GMDPTIS (IN REF. 4).
 FT CONFLICT 276 276 D -> A (IN REF. 4).
 FT CONFLICT 491 491 Y -> H (IN REF. 2 AND 3).
 FT CONFLICT 1057 1057 T -> P (IN REF. 1).
 FT CONFLICT 1110 1112 SGR -> GGP (IN REF. 4).
 SQ SEQUENCE 1816 AA; 201908 MW; 04E9AF379A0F4A4D CRC64;
 Query Match 5.8%; Score 93; DB 1; Length 1816;
 Best Local Similarity 23.8%; Pred. No. 4.2;

DR	Pfam; PF00362; integrin_B; 1.	
DR	PRINTS; PR01186; INTEGRINB.	
DR	ProDom; PD001811; Integrin_B; 1.	
DR	SMART; SM00187; INB; 1.	
DR	SMART; SM00423; PSI; 1.	
DR	SMART; SM00327; VWA; 1.	
DR	PROSITE; PS00243; INTEGRIN_BETA; 3.	
DR	PROSITE; PS00022; EGF_1; UNKNOWN 2.	
KW	Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;	
KW	Repeat; Signal; Phosphorylation.	
FT	SIGNAL	21
FT	CHAIN	22 798
FT	DOMAIN	22 727
FT	TRANSMEM	728 751
FT	DOMAIN	752 798
FT	DOMAIN	139 377
FT	DOMAIN	466 635
FT	REPEAT	466 515
FT	REPEAT	516 559
FT	REPEAT	560 598
FT	REPEAT	599 635
FT	DISULFID	28 464
FT	DISULFID	36 46
FT	DISULFID	39 76
FT	DISULFID	49 65

[illegible]


```
Db 306 NPCTGGMCHSVPLSNTCRCPGYQLDSQVHCVIDECQDSPACDVNTLGSFHC 365
QY 171 CY-----EAKIMSSICKCIACKN 189
Db 366 CWVGQPSGPKKEA---CEDVDECAANS 391

RESULT 13
OC90_HUMAN
ID OC90_HUMAN STANDARD; PRT; 493 AA.
AC Q02509;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Otc90in 90 precursor (Otc90) (Phospholipase A2 homolog).
GN OC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Murthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RT A2 related gene."
RL Nucleic Acids Res. 21:135-143(1993).
CC -!- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z14310; CAA78662.1; ALT_INIT.
DR HSSP; P00593; 4BP2.
DR Genew; HGNC:8100; OC90.
DR MIM; 601658.
DR InterPro; IPR001211; PhospholipaseA2.
DR Pfam; PF00068; phoslip; 2.
DR PRINTS; PR00389; PHPLIPASEA2.
DR ProDom; PD000303; PhospholipaseA2; 2..
DR SMART; SM00085; PA2c; 2.
DR PROSITE; PS00118; PA2_HIS; 2.
DR PROSITE; PS00119; PA2_ASP; 1.
KW Signal; Glycoprotein; Repeat.
FT SIGNAL 1 17
FT CHAIN 18 493
FT DOMAIN 76 190
FT DOMAIN 321 377
FT DOMAIN 389 441
FT DISULFID 85 145
FT DISULFID 99 190
FT DISULFID 101 117
FT DISULFID 116 172
FT DISULFID 123 165
FT DISULFID 132 158
FT DISULFID 152 163
FT CARBOHYD 38 38
FT CARBOHYD 179 179
FT CARBOHYD 423 423
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364A5B7 CRC64;

Query Match 5.7%; Score 91; DB 1; Length 493;
Best Local Similarity 18.3%; Pred. No. 1.5;
Matches 59; Conservative 50; Mismatches 89; Indels 124; Gaps 14;
```

```
QY 2 VICOLKGAQ-MLCIDNCGARELKA-----LHLLPOYDDOSSPQSELPKPMITLVGR 53
Db 150 IICESKDNCEHLLCTCDKAAIECLARSSUNSLNLLDTSCFLAQPTPETTIKEDTLTLPR 209
QY 54 LILPV-PAKLNLI-----TQVD-----NGALPSAVNGAA----- 80
Db 210 VVPVEPTDTSLTALSSEVAATEADRLITLSKKAGHDQEGVCAARATSPPGSAEIVATR 269
QY 81 -----FPSG-----PALQPPKTLTSGYCDCFSSGDFCSCSCNNLRHELERPK 124
Db 270 VTAKIVTLVPAIGIKSLGLAVSSVENDPEETTEKACDRFT---FLHLSGDNMQVMPQLGE 326
QY 125 AIKACILDRNPEAFQPMKMGKRLGAALKRHSKCNCRSG-----CLKNYCECYE 173
Db 327 MLFCLTSRCPEEF-----SYGCGQEGGEGPRDDLRCCLSLHHC-CUE 370
QY 174 -----AKIMCS-----SICKCIACKNYEESPERKM 198
Db 371 QVRRLGCLLERLPWSPVVCVDHTPKCGGSLCKELKLCACDQTAACMTSASFNSQLKSPS 430
QY 199 LMSTHYMEPGDFESSHYLSPA 220
Db 431 RLGCPCG--QPAACEDSLHPVA 450

RESULT 14
EZHL_HUMAN
ID EZHL_HUMAN STANDARD; PRT; 747 AA.
AC Q92800; O43287; Q14459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enhancer of zeste homolog 1 (ENX-2).
GN EZHL OR KIAA0388.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97079676; PubMed=8921387;
RA Abel K.J., Brody L.C., Valdes J.M., Erdos M.R., McKinley D.R.,
RA Castilla L.H., Merajver S.D., Couch F.J., Friedman L.S.,
RA Ostermeyer E.A., Lynch E.D., King M.-C., Welcsh P.L.,
RA Osborne-Lawrence S., Spillman M., Bowcock A.M., Collins F.S.,
RA Weber B.L.;
RT "Characterization of EZHL, a human homolog of Drosophila Enhancer of
RT zeste near BRCAL."
RL Genomics 37:161-171(1996).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=98146265; PubMed=9473645;
RA Ogawa M., Hiraoka Y., Taniguchi K., Aiso S.;
RT "Cloning and expression of a human/mouse Polycomb group gene,
RT ENX-2/Enx-2."
RL Biochim. Biophys. Acta 1395:151-158(1998).
RN [3]
SEQUENCE FROM N.A.
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
RN [4]
SEQUENCE FROM N.A.
RX TISSUE=Brain;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RN [5]
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
```



```

Query Match      5.7%; Score 91; DB 1; Length 2907;
Best Local Similarity 18.3%; Pred. No. 10;
Matches 73; Conservative 36; Mismatches 121; Indels 168; Gaps 21;

QY 4 COLKGGAQ--MLCIDNGCARELKALHLLPQYDDOSSPQSELPKPMTTLVGR----- 53
   | : : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 405 CPVGRSEYRRLCDG-----LP-----MGGIPGSSVSRPGGT--GSGNGGYGPG 447
   | : : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 54 ---LLPVPKAKLNLITVDNGALPSAVNGAAPSPGPAQLQP-----PK 92
   | : : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 448 GTGFLPIPG-----DNGFSP-CVGAGVGAGG--QGPIITGLTILNQITIDICKHHAN 496
   | : : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 93 ITLSGYC-----DCF-----SSGDFCNS-----CSCNNL 116
   | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 497 LCLNGRCITPVSSYRCNMGYKQDANGDCIDVDECTSNPCNSGDCVNTPGSYCKCH-- 554
   | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 117 RHELERPKAKACLDNRPEAFQPKMGKGRGLGAALKRHSKGCNCRKRSGLK----- 166
   | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 555 -AGPQRTPTKQACIDID-ECIQ-----NGVLCKNGRCVNSDGSFQCICN 596
   | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 167 -----NYCEYEAKIMCSSIC-----KCIACKNYEESEPERKMLMSTPHYM 206
   | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db 597 AGFELTWDGKNCDVDEHCTTTNMLCNGMCINEDGSFCVKCPGPFILAPNGRYCTVDDECQ 656
   | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 207 EPGDFESSHYLS-----PAKFSGPPLKRNKRAFSCIS-----WEVVEATCA----- 248
   | : | | | : | | : | | : | | : | | : | | : | | : | | : | | : |

```

Db 657 TPGICMNGHCINNEGSRFCDCPPGLAVGVDGRVCVDTNMRSTCYGEIKKGVCVRPFPGAV 716
Qy 249 ----CLLAOGEEAEQEHCSPLAEQMILEEFGRLSQI 282
Db 717 TKSECCCANPDYGRGEPQCPAKNS-AEFHGLCSSGI 753

Search completed: April 21, 2003, 11:48:56
Job time : 12.9394 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:30:23 ; Search time 32.7778 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLIHTEFKSLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	100.0	295	21	Mouse testis speci
2	1251	78.2	299	21	Human testis speci
3	435	27.2	950	22	Drosophila melanog
4	396	24.8	438	20	Caenorhabditis ele
5	362	22.6	280	22	Novel human diagn
6	322	20.1	147	22	Human testicular a
7	322	20.1	147	22	Human reproductive
8	248	15.5	53	22	Human colon cancer
9	186	11.6	243	22	Drosophila melanog
10	165	10.3	251	22	Human polypeptide,

11	110.5	6.9	732	18	AAW26642	Human RECK cancer-
12	99	6.2	2931	22	ABB68229	Drosophila melanog
13	97	6.1	602	21	AAV95660	Human Zntfr2 protel
14	97	6.1	1987	22	ABB61083	Drosophila melanog
15	96.5	6.0	359	21	AAG09497	Arabidopsis thalia
16	96.5	6.0	443	21	AAG09496	Arabidopsis thalia
17	96.5	6.0	472	21	AAG09495	Arabidopsis thalia
18	96.5	6.0	1788	22	ABB62995	Drosophila melanog
19	96.5	6.0	4547	22	ABB59051	Drosophila melanog
20	96	6.0	773	22	ABB59751	Drosophila melanog
21	96	6.0	3635	23	ABB81589	Mouse laminin alph
22	96	6.0	3635	23	AAW50357	Mouse laminin-15 a
23	95.5	6.0	1035	22	ABB66062	Drosophila melanog
24	94.5	5.9	759	23	AAU97037	Human LP protein L
25	94	5.9	1792	21	ABA48443	Human laminin 8 po
26	94	5.9	1800	21	ABA48445	Human laminin 8 po
27	94	5.9	1816	21	ABA48442	Human laminin 8 po
28	94	5.9	1824	21	ABA48444	Human laminin 8 po
29	93.5	5.8	652	19	AAW49879	Amino acid sequenc
30	93.5	5.8	652	21	AAW32345	Human cell surface
31	93.5	5.8	1981	22	ABB61657	Drosophila melanog
32	92.5	5.8	921	21	AAV32297	Corn polycomb prot
33	92	5.8	3680	22	ABB70878	Drosophila melanog
34	91.5	5.7	637	21	AAV32346	Mouse cell surface
35	91.5	5.7	644	21	AAV79186	Haematopoietic ste
36	91.5	5.7	644	21	AAV79193	Haematopoietic ste
37	91	5.7	18	21	AAV68465	Mouse testis speci
38	90	5.6	533	23	ABB53284	Human polypeptide
39	89.5	5.6	347	22	AAW87443	Human gene 35 enco
40	89.5	5.6	361	22	AAW99918	Human polypeptide
41	89.5	5.6	439	22	AAW95609	Human protein sequ
42	89	5.6	1725	21	AAW19800	Mouse laminin 2 ma
43	89	5.6	1725	21	ABA48451	Mouse laminin 10 s
44	89	5.6	1725	23	ABB81593	Mouse laminin 10 s
45	89	5.6	1764	10	AAW91672	Primary amino acid

ALIGNMENTS

RESULT 1

AAV68463

ID AAV68463 standard; Protein; 295 AA.

XX AAV68463;

AC AAV68463;

DT 25-APR-2000 (first entry)

XX Mouse testis specific factor tesmin SEQ ID NO:4.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
XX differentiation regulatory factor; male germ cell regulatory actor;
XX germ cell differentiation; sterility.

OS Mus musculus.

XX WO200004147-A1.

PN 27-JAN-2000.

PD 16-JUL-1999; 99WO-JP03859.

PF 17-JUL-1998; 98JP-0219856.

PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI; 2000-147785/13.

DR N-PSDB; AAZ88155, AAZ88157.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility
XX
PS Claim 1; Page 47-49; 63pp; Japanese.
XX
CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 295 AA;
Query Match 100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 3e-144;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPPKMTLVGRLLPVPK 60
DB 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPPKMTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPIKTLGSGYCDGFCSSGDFCNSC 120
DB 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPIKTLGSGYCDGFCSSGDFCNSC 120
QY 121 ERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQ 180
DB 121 ERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQ 180
QY 181 ICCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAF 240
DB 181 ICCKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAF 240
QY 241 EVVEATCACLQAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
DB 241 EVVEATCACLQAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
RESULT 2
AAY68464
ID AAY68464 standard; Protein; 299 AA.
XX
AC AAY68464;
XX
XX
DT 25-APR-2000 (first entry)
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
XX Homo sapiens.
XX
XX WO200004147-A1.
XX
XX 27-JAN-2000.
XX
XX 16-JUL-1999; 99WO-JP03859.
XX
XX 17-JUL-1998; 98JP-0219856.
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX
XX WPI; 2000-147785/13.
XX N-PSDB; AA288156.
XX
XX New male germ cell regulatory factor tesmin expressed in spermatocytes
XX useful for investigation of germ cell differentiation and sterility
XX
XX Claim 1; Page 50-52; 63pp; Japanese.
XX

CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin. Tesmin
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 299 AA;
Query Match 78.2%; Score 1251; DB 21; Length 299;
Best Local Similarity 76.3%; Pred. No. 5.1e-111;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;
QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPPKMTLVGRLLPVPK 60
DB 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPPKMTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPIKTLGSGYCDGFCSSGDFCNSC 116
DB 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPIKTLGSGYCDGFCSSGDFCNSC 120
QY 117 RHELEKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQ 176
DB 121 HHDIERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQ 180
QY 177 MCSSICKICGCKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAF 236
DB 181 MCSSICKICGCKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAF 240
QY 237 CISWEVVEATCACLQAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
DB 241 CISWEVVEATCACLQAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 299
RESULT 3
ABB62035
ID ABB62035 standard; Protein; 950 AA.
XX
AC ABB62035;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 12897.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL06138.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions
XX
XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of

CC Insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 950 AA;

Query Match 27.2%; Score 435; DB 22; Length 950;
 Best Local Similarity 38.4%; Pred. No. 1.7e-32;
 Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

QY 31 QYDQSQFFQSELPKPTTLVGRLLPVPKLNLTQVDNGALPSAVNGAAPSPGALQGP 90
 DB 681 QLPTEQSTPIKVEKLTPLPGVKANVPK--PLFEVLKPPATAAAGAVDPLGGMTSRR 738
 QY 91 PKITLSG-----YDCDFSSGDFCNSCSC-----NNLRHELERFKAIKACLDNRNPEAFQPK 140
 DB 739 KHCNCKSQCLKLYCFCFANGFCQCTCKDCFNLDYEVERAIRSCLDNRNPSAFKPK 798
 QY 141 MGKRLGAALKRHSKGNCRSGCLKNYCEYEAKEIMCCKICKIACKNYEESPERKMLM 200
 DB 799 ITAPNSGDMRL-HNKGNCNCRSGCLKNYCEYEAKEIPCSSICKCVGCRNMDRPD----- 852
 QY 201 STPHMEPGDPFESSHYLSPAKFSGPPKLRNRO-----AFSCISWEVVEATCACLLAQGE 255
 DB 853 -----VDMSDLGLMGVGGQKDK-ARNKQLNENRANIYFTDDVIEATIMCISRIV 903
 QY 256 EAEQHCSPSLAEQMIIEFGRCILSQIL 283
 DB 904 MHEQKNVAVEDMEREVMEEMGESLTQII 931

RESULT 4

AAW83392
 ID AAW83392 standard; Protein; 438 AA.

XX
 AC AAW83392;

XX 29-MAR-1999 (first entry)

XX Caenorhabditis elegans synMuv protein LIN-54.

XX LIN-54; synthetic multivulvar; SynMuv; signal transduction;
 KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 KW cell proliferation; gene therapy.

XX Caenorhabditis elegans.

XX WO9854299-A1.

XX 03-DEC-1998.

XX 28-MAY-1998; 98WO-US11043.

XX 28-MAY-1997; 97US-0047996.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Ceol C, Horvitz HR, Lu X;

XX WPI; 1999-045362/04.

XX N-PSDB; AAV72865.

XX Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas

XX Claim 7; Fig 13; 70pp; English.

XX This is the amino acid sequence of LIN-54, a novel protein of

CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
 CC proliferation, and is part of a pathway that may be used as a
 CC genetic and biochemical model system for tumour suppression and
 CC cancer in mammals. SynMuv pathway genes and proteins may be used
 CC to identify genes which are part of the mammalian pathway and
 CC to identify genes, proteins and therapeutic compounds which
 CC modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 CC C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 CC -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 CC acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 CC polypeptide, and (5) an antibody which binds to a SynMuv family
 CC protein. The SynMuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions
 CC involving altered levels of cell proliferation, e.g. SynMuv-
 CC associated carcinomas.

XX
 SQ Sequence 438 AA;

Query Match 24.8%; Score 396; DB 20; Length 438;

Best Local Similarity 38.9%; Pred. No. 3.4e-29;

Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

QY 98 YDCDFSSGDFCNSCSC-----NNLRHELERFKAIKACLDNRNPEAFQPKMGKRLGAALKR- 152
 DB 191 YDCDFANGFCRCNCKDCHNNIETDSQSKAIRQSLERNPNAPKPKIGIARGGTIDIER 250
 QY 153 -HSKGCNCRSGCLKNYCEYEAKEIMCCKICKIACKN-----YVES-----PERKML 199
 DB 251 LHQSGCHCKKSGCLKNYCEYEAKEVPCTDRCKCKGCQNTETRYMTRYKNSGGAVSNTNAL 310
 QY 200 MS-----TPHYMEPGDPFESSHY-----LSPAKFSGPPKLRNROAFSCISN-----EVV 243
 DB 311 MSFTNASSTATPD-SGPGSVVTVDEHGDYEDMLLSHKPKVEMDPRPF---PWYMTDDEV 366
 QY 244 EATCACLAAQGEA-----BOEHCSPLSABQMILEEFGRCILSQIL 283
 DB 367 EATMCMVAQAEALNYEKVQTEDEKLNMEKLVLEFGRCLEQMI 412

RESULT 5

ABG17958
 ID ABG17958 standard; Protein; 280 AA.

XX
 AC ABG17958;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #17949.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS82145.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer
PT
XX
PS Claim 11; SEQ ID NO 1409; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and

CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a protein of the
CC invention.
XX
SQ Sequence 147 AA;
Query Match 20.1%; Score 322; DB 22; Length 147;
Best Local Similarity 73.8%; Pred. No. 9.7e-23;
Matches 62; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
QY 1 MVICQLKGGAGMCLDCGARELKLHLLPQYDDQSFQSELPKMTTLVGRLLPVPK 60
Db 60 MVICQLKGGTQMLCIDNSRTELKALHLLPQYDDQNNYLSQDVPKMTALVGRFLPASTK 119
QY 61 LNLITQVDNGALPSAVNGAAPPSSG 84
Db 120 LNLITQLEGALPSVYNGSAPPSSG 143
RESULT 7
AAM95330
ID AAM95330 standard; Protein; 147 AA.
XX
AC AAM95330;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 3988.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465570/50.

DR N-PSDB; AAL01300.

XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.

XX SQ Sequence 147 AA;

Query Match 20.1%; Score 322; DB 22; Length 147;

Best Local Similarity 73.8%; Pred. No. 9.7e-23;

Matches 62; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGRLLPVPK 60

Db 60 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGRLLPVPK 119

QY 61 LNLITVDNGALPSAVNGAAPPSG 84

Db 120 LNLITQLEALPSVWNGSAPPSSG 143

RESULT 8

ID AAG76158 standard; Protein; 53 AA.

XX AAG76158;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6922.

DE

XX

XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 95JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX N-PSDB; AAK94268.
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX Sequence 251 AA;
Query Match 10.3%; Score 165; DB 22; Length 251;
Best Local Similarity 75.6%; Pred. No. 1.9e-07;
Matches 31; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
Qy 1 MVICOLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLQS 250
Db 210 MVICOLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLQS 250
RESULT 11
ID AAW26642
XX AAW26642 standard; Protein; 732 AA.
XX AC AAW26642;
XX 11-FEB-1998 (first entry)
XX Human RECK cancer-inhibiting protein.
XX RECK; reversion-inducing cysteine rich protein with Kazal motif;
XX human; cancer; gene therapy; diagnosis.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Domain 396..415
XX /label= Kazal_domain
XX WO9724439-A1.
XX 10-JUL-1997.
XX 24-DEC-1996; 96WO-US20812.
XX 27-DEC-1995; 95JP-0340469.
XX (AMGE-) AMGEN INC.
XX (KITA/) KITAYAMA H.

PA (NODA/) NODA M.
PA (TAKA/) TAKAHASHI C.
PA (SANY) SANKYO CO LTD.
XX Kitayama H, Noda M, Takahashi C;
XX WPI; 1997-363675/33.
XX N-PSDB; AAT90508.
XX RECK gene and corresponding protein sequences - enables reversion of
XX cancer cells
XX Claim 7; Page 42-44; 53pp; English.
XX This protein sequence comprises the human RECK protein (reversion-
XX inducing cysteine rich protein with Kazal motif), which is capable
XX of transforming malignant cancer cells with an activated ras gene
XX into normal cells (reversion activity). Its amino acid sequence
XX was deduced from a cDNA clone (see AAT90508) obtained from human
XX fibroblast MRC-5 (ATCC CCL 171) cells. A claimed method for
XX treatment of cancer comprises contacting the cancer cells with a
XX RECK polypeptide. RECK may also be expressed using gene therapy
XX methods for in vivo treatment of cancer.
XX Sequence 732 AA;
Query Match 6.9%; Score 110.5; DB 18; Length 732;
Best Local Similarity 17.9%; Pred. No. 0.12;
Matches 63; Conservative 41; Mismatches 110; Indels 137; Gaps 17;
Qy 4 COLKGGAMLCDNCGARELKALHLLPOYDDOSSPQSELKPKMTLVG-----RLLPV 57
Db 183 CHSKSRGSIICKSDC-----VEILKKCGDQNKFPEDHTAESICELLSPTDLKNCIPL 235
Qy 58 -----PAKL-NLITQVDNALPS-----AVNGAARPSG-PALQGPKITLSGYCDGF 102
Db 236 DTYLRPSTLGNIVEEVTHPCNPNCANELCEVNRKKGCPGDPCLI---PYFCVQG-CKLG 291
Qy 103 SSGDFCNSCSCNNLRHELEKAKACLDNRNPEAFQPKMKGRLGAALRHSKGCNCKRS 162
Db 292 EASDF-----IVROGTLIQVPSSAGAGEVCYKI-----CSCQGS 324
Qy 163 GCLK-----NYCEYEAKIMCSS----- 180
Db 325 GLLNCMEMHCHDLQKSCIVGGRKRSHTSFSIDCNVSCFAGNLVCTRLCLSEHSSD 384
Qy 181 -----ICKCI-----AC-----KNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFS 223
Db 385 DRRTFTGLPCADQFVPCVQNGRTYPSACIARCVGLQDHQDFEGSCMSK---DPC--- 438
Qy 224 GPPKLRKNRQA-----FSCISWEVVEATCACLIAQGEARQEH 261
Db 439 NPNCKQNRQICPKPQVCLTTFDKFGCSQYECVPRQLACDQVQDPVCDTH 489
RESULT 12
ID ABB68229
XX ABB68229 standard; Protein; 2931 AA.
XX AC ABB68229;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 31479.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.
XX PF
XX 23-MAR-2000; 2000US-191637P.
XX PR
XX 11-JUL-2000; 2000US-0614150.
XX PA
XX (PEKE) PE CORP NY.
XX PI
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658860/75.
XX DR N-PSDB; ABL12332.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions
XX
XX Disclosure; SEQ ID NO 31479; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX Sequence 2931 AA;

Query Match 6.2%; Score 99; DB 22; Length 2931;
Best Local Similarity 21.3%; Pred. NO. 8.6;
Matches 50; Conservative 26; Mismatches 81; Indels 78; Gaps 11;

QY 100 DCFSSGDFCNCS-----CNLRLHLERFKAKACLDNRNPAFQPK 140
Db 1677 DCASSGGLTKCGYDNCVTFVSDSGSVTQRCSE-----SVFEESDYC-DENP-AYCPR 1729

QY 141 MCKGRGAALRHSKGCNKRSGCLKNYCEYKIMCISIC-----KCI 185
Db 1730 C-----NSNGCNTADS--QDKYVEICVDSSVDSNCVSDPTQITKTROCHERC 1776
QY 186 AC-----KNYESPERKMLMSTHYMEPGDFESSHYLSPAKFGSPKLRKNRQAFSCISWE 241
Db 1777 SAFPLPNETEDPSYALIRNCYDLEKEDRD-----ACTAGSKRFCATCGTGKCNSED 1829

QY 242 VYEATCACLQAQEEA-----EQEHCPSLAFQOMILEEFGRLCSOLH 284
Db 1830 LVASRHSCLVCGDECSQSQASSCSNYRHHDECYIOFDEERSITSLG-CLSELH 1883

RESULT 13
AAV95660
ID AAV95660 standard; Protein: 602 AA.
XX AC AAV95660;
XX 25-OCT-2000 (first entry)
XX Human Znt2 protein.
XX Znt2; epidermal growth factor-like domain; human;
KW cell differentiation; vulnery; diagnosis; therapy.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..30
FT /label= Signal_peptide
FT /note= "the signal peptide alternatively

FT Protein 31..602 comprises amino acid residues 9..30"
FT Domain 31..507
FT /note= "extracellular domain"
FT Domain 508..533
FT /note= "transmembrane domain"
FT Domain 534602
FT /note= "intracellular domain"
FT Region 224..256
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 224..254 or 224..251"
FT Region 272..303
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 272..301 or 272..298"
FT Region 317..351
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..349 or 317..346"
FT Region 371..402
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..400 or 317..397"
FT Region 422..449
FT /note= "laminin-type epidermal growth factor-like
FT domain signature"
XX WO200043512-A1.
XX 27-JUL-2000.
XX
XX 20-JAN-2000; 2000WO-US01419.
XX
XX 25-JAN-1999; 99US-0237074.
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Holloway JL, Lofton-Day CE, Gilbert T;
XX WPI; 2000-491163/43.
XX N-PSDB; AAA50101, AAA50102.
XX
XX Isolated Znt2 nucleic acids and polypeptides which act as epidermal
XX growth factors, useful for the treatment of e.g. kidney and liver
XX disorders, burns, and ulcers and for regulating smooth muscle cell
XX proliferation
XX
XX Claim 1; Page 89-91; 98pp; English.
XX
XX The present sequence is of that of human Znt2, a new member of the
XX epidermal growth factor-like domain group of proteins. The
XX sequence was deduced from isolated Znt2 cDNA (see AAA50101). In
XX addition to the presence sequence, polypeptides comprising amino
XX acids 31-507 and 31-602 of the sequence are claimed. An expression
XX vector, a recombinant host cell, a method of using the vector to
XX produce Znt2 protein, an antibody or antibody fragment that
XX specifically binds to the polypeptide, and a method of detecting
XX the presence of Znt2 gene products in a biological sample are also
XX all claimed. Znt2 polypeptides may be used to regulate vascular
XX smooth muscle cell proliferation, to restore normal neurological
XX functioning after trauma, to treat ocular disorders, to treat
XX kidney and liver disorders, to promote hair and follicular
XX development, to stimulate growth and differentiation of various
XX epidermal and epithelial cells in vivo and in vitro, for the
XX treatment of burns, ulcers and corneal incisions, and to stimulate
XX wound healing.
XX
XX Sequence 602 AA;

Query Match 6.1%; Score 97; DB 21; Length 602;
Best Local Similarity 21.8%; Pred. No. 1.8;
Matches 56; Conservative 26; Mismatches 101; Indels 74; Gaps 16;

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:48:39 ; Search time 13.4091 Seconds
(without alignments)
647.304 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGAQMLCIDNCGA.....GRCLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93.5	5.8	652	US-08-751-305-2	Sequence 2, Appli
2	87.5	5.5	610	US-08-365-470-3	Sequence 3, Appli
3	87.5	5.5	610	US-09-209-668-19	Sequence 19, Appl
4	87.5	5.5	610	US-09-009-490A-89	Sequence 89, Appl
5	86	5.4	418	US-08-795-430-13	Sequence 13, Appl
6	86	5.4	418	US-09-355-700-13	Sequence 13, Appl
7	86	5.4	610	5217870-2	Patent No. 5217870
8	86	5.4	830	US-08-110-158-4	Sequence 4, Appli
9	85.5	5.3	769	US-08-789-078-1	Sequence 1, Appli
10	85.5	5.3	769	US-08-752-633-1	Sequence 1, Appli
11	85.5	5.3	769	US-08-476-062A-45	Sequence 45, Appl
12	85.5	5.3	769	US-07-728-215-31	Sequence 31, Appl
13	85.5	5.3	769	US-08-938-085A-31	Sequence 31, Appl
14	85.5	5.3	769	PCT-US95-04886-1	Sequence 1, Appli
15	85.5	5.3	769	PCT-US96-01314-45	Sequence 45, Appl
16	85	5.3	217	US-09-602-543-5	Sequence 5, Appli
17	85	5.3	235	US-09-602-543-4	Sequence 4, Appli
18	85	5.3	676	US-08-630-172-10	Sequence 10, Appl
19	85	5.3	676	US-09-375-419-10	Sequence 10, Appl
20	84.5	5.3	696	US-08-899-437-23	Sequence 23, Appl
21	84.5	5.3	696	US-09-126-121-23	Sequence 23, Appl
22	84.5	5.3	720	US-08-899-437-6	Sequence 6, Appli
23	84.5	5.3	720	US-09-126-121-6	Sequence 6, Appli
24	83.5	5.2	350	US-08-999-811-4	Sequence 4, Appli
25	83.5	5.2	350	US-08-824-996-2	Sequence 2, Appli
26	83.5	5.2	350	US-09-042-105-4	Sequence 4, Appli
27	83.5	5.2	350	US-08-510-133A-33	Sequence 33, Appl

28	83.5	5.2	350	4	US-08-585-895-33	Sequence 33, Appl
29	83.5	5.2	419	2	US-08-999-811-2	Sequence 2, Appli
30	83.5	5.2	419	3	US-09-042-105-2	Sequence 2, Appli
31	83.5	5.2	419	3	US-09-042-105-18	Sequence 18, Appl
32	83.5	5.2	419	4	US-08-795-430-8	Sequence 8, Appli
33	83.5	5.2	419	4	US-08-510-133A-35	Sequence 35, Appl
34	83.5	5.2	419	4	US-09-355-700-8	Sequence 8, Appli
35	83.5	5.2	419	4	US-09-355-700-58	Sequence 58, Appl
36	83.5	5.2	419	4	US-08-601-132-33	Sequence 33, Appl
37	83.5	5.2	419	4	US-08-706-054A-3	Sequence 3, Appli
38	83.5	5.2	419	5	PCT-US96-09001-2	Sequence 2, Appli
39	81.5	5.1	1964	4	US-09-467-997-1	Sequence 1, Appli
40	81	5.1	415	4	US-08-795-430-11	Sequence 11, Appl
41	81	5.1	415	4	US-09-355-700-11	Sequence 11, Appl
42	81	5.1	909	4	US-09-013-895A-4	Sequence 4, Appli
43	81	5.1	909	4	US-09-448-868-4	Sequence 4, Appli
44	81	5.1	2476	2	US-08-276-967-2	Sequence 2, Appli
45	80.5	5.0	3224	2	US-08-705-660-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Query Match 5.8%; Score 93.5; DB 2; Length 652;
Best Local Similarity 27.0%; Pred. No. 0.66;
Matches 30; Conservative 11; Mismatches 51; Indels 19; Gaps 4;
QY 96 SGVC--DCFFSGDFCNSCNNLRHLERFKAKICLDNRPEAFQPKMGKRLGAALRH 153
Db 268 NGCHQDCEGGDGSFLCGC---RPGFRLLDDLVTCASRNPCSSSPGRCG----- 313
QY 154 SKGCNCRSGCKLKNY-CECYEAKIMC-SSICKICIAKNYEESEPKMLMSTP 203
Db 314 --GATCVLPHGKNYTCRCPOGVQLDSSQLDCVDVDECQDSPCAQECVNTP 362

```

RESULT 2
US-08-365-470-3
: Sequence 3, Application US/08365470
: Patent No. 5632991
: GENERAL INFORMATION:
: APPLICANT: Gimbrone, Jr., Michael A.
: TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
: TITLE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
: STREET: 1100 New York Ave., NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release. #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/365,470
: FILING DATE: herewith
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/102,510
: FILING DATE: 05-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/850,802
: FILING DATE: 13-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Markowicz, Karen R.
: REGISTRATION NUMBER: 36,351
: REFERENCE/DOCKET NUMBER: 0627.1350003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 610 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-365-470-3

```

```

Query Match      5.5%; Score 87.5; DB 1; Length 610;
Best Local Similarity 21.6%; Pred. No. 2.4;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNGARE-----LKAHLHLLPOYDDOS----SFOSELPKPMITL-----VGR-LLPVPA 59
DB 177 IVNCTALESPEHGSILVCSHPLGNFSYNSSCISCDRGYLPSSMETMQCMSSGSAWAPPA 236
QY 60 ----KLNLITQVDNALPSAVNGAAPP-----SGPALOGPKKITLSGYCDCFSWG 105
DB 237 CNVVECDAVTNPANGFVECFQNGPSPFWNTTCTFDCEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSNNLRHLELEKAKACLDLRNPEAFOPKMGKGRLG---AAKLHRSKG 156
DB 291 NWDNEKPTCAVTCRAVR-----QPONGSVRCSHPAGEFTTKSS 330
QY 157 CN--CKRSGCLKNYCEYEAKIMGSS-----ICKIACKNYEESPERKMLMSTPHY 205
DB 331 CNFTCEEGFMLOG-----PAQVECTTQGWTOQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME--PGDFESSHYLSPAKFSPPP-----KLKRNRAQFCSISHEVVEATCACLLAQ----- 253
DB 378 MNCLPSASGSRFYSGSSCEFFSCQEGFVLKGRKRCOGPTGEWMDNEKPTCEAVRCRDVHQPP 437
QY 254 -----GEAEQEHCSPLAEQMILEEFQ-----RCLSQ 281

```

Db 438 KGLVCAHSPICEFTYKSSCAFSCEEG--FELYGSTOLECTSQ 478

```

RESULT 3
US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

```

Query Match	5.5%	Score 87.5;	DB 3;	Length 610;
Best Local Similarity	21.6%;	Pred. No. 2.4;		
Matches	74;	Conservative 33;	Mismatches 119;	Indels 117;
Gaps				
QY	15	IDNCGARE-----LKALHLLPYDDQS-----SFQOSELPKPMNTLL-----VGR-LLPVPA 59		
Db	177	IVNCTALESPHEGSLVCVSHPLGNFYSNCSISCDRGYLPSSMETMQCHSSGEWSAPIPA 236		
QY	60	-----KMLNLTQVQDNGALPSAVNGAAPP-----SGPALQGPPIKTLTSGYCDCFSSG 105		
Db	237	CNVVECDAVTPANGVECFQNGPSFPWNVTCTFDCEEGFELMGAQSL-----QCTSSG 290		
QY	106	DF-----CNSCSNNLRHELRFAKIAKACLDNPEAFQPMKGKRLG---AKLHRSG 156		
Db	291	NWDNEKPTCKAVTCAVR-----OPQNGSVRCSHSPAGEFTFKSS 330		
QY	157	CN--CKRSGCLKNVCEYEAKIMCSS-----ICKCIACKNYEESPERKMLNSTPHY 205		
Db	331	CNFTCEGFMLOG-----PAQVECTTQGWTOQIYCEAFQCTAL--SNPER-----GY 377		
QY	206	ME--PGDFESSHYLSPAKFSGPP-----KLKKNQAFSCISWEVVEATCACLLAQ-----253		
Db	378	MNCLPSNCGSFRYGSCEFSCEBGFVLKSKRLQCGPTGWDNEKPTCEAVRCDAVHQPP 437		
QY	254	-----GGEAEQHCSPSLAEQMILEEFG-----RCLSQ 281		
Db	438	KGIVRCAHSPIGFTFYKSSCAFSCEBG--FELYGSTOLECTSO 478		

```

RESULT 4
US-09-009-490A-89
; Sequence 89, Application US/090009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WORDPERFECT 6.0

```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/009,490A
;; FILING DATE: January 20, 1998
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 440,740
;; FILING DATE: May 12, 1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 063,167
;; FILING DATE: May 17, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 969,151
;; FILING DATE: February 10, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 007,997
;; FILING DATE: January 20, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 939,855
;; FILING DATE: September 2, 1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 567,286
;; FILING DATE: August 14, 1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jane Massey Licata
;; REGISTRATION NUMBER: 32,257
;; REFERENCE/DOCKET NUMBER: ISPH-0268
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (609) 810-1515
;; TELEFAX: (609) 810-1454
;; INFORMATION FOR SEQ ID NO: 89:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 610
;; TYPE: Amino Acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;; ANTI-SENSE: no
US-09-009-490A-89

Query Match 5.5%; Score 87.5; DB 4; Length 610;
Best Local Similarity 21.6%; Pred. No. 2.4; Indels 117; Gaps 20;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;
QY 15 IDNCGARE-----LKAHLHPQYDDQS-----SFPQSELPKPMTTL-----VGR-LLPVPA 59
Db 177 IVNCTALESPEHSLVCSHPLGNFNSNCSISCDRGYLPSSMETMQCMSSGSEWSAPIPA 236
QY 60 ----KLLITQVDNGALPSAVNGAAPP-----SGPALQGGPKITLGYDCDFSSG 105
Db 237 CNVVECDVATNPANGFVECFQNPGSFPWNTTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLHELEFRKAIKACLDNRNPEAFQPKMGKGRLG---AAKLHRSKG 156
Db 291 NMDNEKPTCKAVTCRAVR-----OPQNGSVRCSSHPAGEFTFKSS 330
QY 157 CN--CKRSCLKNYCEYEAKIMCSS-----ICKIACKNVEESPERKMLMNSTPHY 205
Db 331 CNFTCEGFMLOG-----PAQVECTQGTQTOIPVCEAFQCTAL-SNFER-----CY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLRKNRQAFSCISWEVVEATCACLIAQ-----253
Db 378 MNCLPSASGSFRYSGSCFSCBQGFVLKGSKRLQCGPTCEWDNEKPTCEAVRCDVHQPP 437
QY 254 -----GEEAEQHCFSLSLAQMLIEFG-----RCLSQ 281
Db 438 KGLVRAHSPIGFTYKSSCAFSCEEG---FELYGSTQLECTSQ 478

RESULT 5
US-08-795-430-13
; Sequence 13, Application US/08795430
; Patent No. 6130071
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari

;; APPLICANT: Joukov, Vladimir
;; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
;; NUMBER OF SEQUENCES: 57
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/795,430
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/FI96/00427
;; FILING DATE: 01-AUG-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/671,573
;; FILING DATE: 28-JUN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/601,132
;; FILING DATE: 14-FEB-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/585,895
;; FILING DATE: 12-JAN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/510,133
;; FILING DATE: 01-AUG-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/340,011
;; FILING DATE: 14-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gass, David A.
;; REGISTRATION NUMBER: 38,153
;; REFERENCE/DOCKET NUMBER: 28967/33691
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 13:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 418 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-795-430-13

Query Match 5.4%; Score 86; DB 4; Length 418;
Best Local Similarity 18.6%; Pred. No. 2;
Matches 48; Conservative 24; Mismatches 90; Indels 96; Gaps 10;
QY 14 CIDNCGARELKAHLHPQYDDQSSF-----POSELPKPM-----48
Db 161 CGCGCNSEGLQCMNISTNYISKTFLFTVPLSHGPKPTVTSFANHSTSCRCMSKLDYVRQV 220
QY 49 -TLVGRLLPVPK-----LNLTQVDNGALPSAVNGAAPPSPALQ 88
Db 221 HSIIRSLPATQTCQHVANKTCPKNHVWNNQICRCLAQHDG-FSSHGLGSDTSEGFHC 279
QY 89 GPKKITLGYCFCDFSSGDF-----CN-----SCSC-----NNL-----116
Db 280 GPNKELDETCQCVCKGGVRRPISCGPHKELDRASCOCMKKLLPSSCGPNKEFDEKQ 339
QY 117 -----RHELEFRKAIKACLDNRNPEAFQPKMGKGRIGLGAALKRHSKGCNCKRSGLKN 167
Db 340 CVCKKTCPKHHPLNPAKCICECTESPNKCFL-----KG-----KRFHHQTCSCYRPCTVR 390

QY 168 YCEYEAKIMCSSICKCI 185
Db 391 TKRCDAGFLLAEEVCRVC 408

RESULT 6
US-09-355-700-13
; Sequence 13, Application US/09355700
; Patent No. 6361946
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; Helmsink University Licensing
; Alitalo, Kari (U.S. only)
; Jonkov, Vladimir (U.S. only)
; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
; Protein and Gene, Mutants Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/355,700
; FILING DATE: 05-NOV-1994
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/795,430
; FILING DATE: 05-FEB-1997
; APPLICATION NUMBER: PCT/F196/00427
; FILING DATE: 01-AUG-1996
; APPLICATION NUMBER: 08/671,573
; FILING DATE: 28-JUN-1996
; APPLICATION NUMBER: 08/601,132
; FILING DATE: 14-FEB-1996
; APPLICATION NUMBER: 08/585,895
; FILING DATE: 12-JAN-1996
; APPLICATION NUMBER: 08/510,133
; FILING DATE: 01-AUG-1995
; APPLICATION NUMBER: 08/340,011
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 28967/34140
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-355-700-13

Query Match 5.4%; Score 86; DB 4; Length 418;
Best Local Similarity 18.6%; Pred. No. 2;
Matches 48; Conservative 24; Mismatches 90; Indels 96; Gaps 10;
QY 14 CIDNCGARELKHLLPOYDQSSF-----PQSELKPMTL----- 48
Db 161 CGGCCNSEGLOCMNISTNYISKTLEITVPLSHGPKPTVTSFANHNTSCRCMKLDVYRQV 220

QY 49 -TLVGRLLPVPK-----LNLITQVDNGALPSAVNGAAPPSPALQ 88
Db 221 HSIIRSLPATQTOCHVANKTCPKNHNWNNQICRLAQHDFG-FSSHLGSDSIDSEGHIC 279

QY 89 GPKRITLSGYCDCFSSGDF-----CN-----SCSC-----NNL----- 116
Db 280 GPNKELDEETQCQVCKGVRPISCGPHKELDRASCQCMCKNKLPLSSCGPNKEFDEKQC 339

QY 117 -----RHELEFRKAIKACLDNRNPEAFQPKMGKRLGAALKRHSKGCNKRSGGLKN 167
Db 340 CVCKKCTCPKHPLNPAKICICECTESPKNCEL-----KG-----KRFHHQTCSYRPPCTVR 390

QY 168 YCEYEAKIMCSSICKCI 185
Db 391 TKRCDAGFLLAEEVCRVC 408

RESULT 7
5217870-2
; Patent No. 5217870
; APPLICANT: HESSON, CATHERINE A.; LOBB, ROY R.; GOELZ, SUSAN E.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST CDX
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/345,151
; FILING DATE: 28-APR-1989
; SEQ ID NO: 2;
; LENGTH: 610
5217870-2

Query Match 5.4%; Score 86; DB 6; Length 610;
Best Local Similarity 22.2%; Pred. No. 3.3;
Matches 64; Conservative 29; Mismatches 101; Indels 94; Gaps 17;
QY 15 IDNCGARE-----LKALHLLPOYDQSS-----SPPQSELKPMTL-----VGR-LLPVA 59
Db 177 IVNCTALESPEHSGSLVCSHPLGNFNSYNSCSISCDRGYLPSSMETMQCMSSGWSADIPA 236

QY 60 ----KLNLIQVDNGALPSAVNGAAPP-----SGPALQPPKATILSGYDCDFSSG 105
Db 237 CNVVECDVNTNANGFVECFNFGFPWNTTCTDCBEGFELMGAQSL-----QCTSSG 290

QY 106 DF-----CNSCSNNLRHELEFRKAIKACLDNRNPEAFQPKMGKRLG---AAKLHSHK 156
Db 291 NWDNEKPTCKAVTCRAVR-----QPONGSVRCSHSPAGEFTFKSS 330

QY 157 CN--CKRSGCLKNYCEYEAKIMCSS-----ICKCIACKNYEESPCKMLMSTPHY 205
Db 331 CNFTCEBGFMLQ-----PAQVECTQGWTOQIPVCEAFQCTAL-SNPER-----GY 377

QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATC 247
Db 378 MNCLPSASGSGFRYSGSCFESCEQGVLUKSKRLQCCGPTGEMDNEKPTC 425

RESULT 8
US-08-110-158-4
; Sequence 4, Application US/08110158
; Patent No. 5603821
; GENERAL INFORMATION:
; APPLICANT: McEver, Rodger P.
; APPLICANT: Pan, Junliang
; TITLE OF INVENTION: Expression Control Sequences of the
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 17
; ADDRESS: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110.158
; FILING DATE: 19930820
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320,408
; FILING DATE: 08-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-815-6508
; TELEFAX: (404)-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-110-158-4

Query Match 5.4%; Score 86; DB 1; Length 830;
Best Local Similarity 22.4%; Pred. No. 5;
Matches 46; Conservative 24; Mismatches 79; Indels 56; Gaps 12;

QY 84 GPALOGPKITLGYDCDFSGDF-----CNSCCNNLRHELEKFKAIKACLDNRPEAF 137
Db 233 QYQVNGPSKL-----ECLASGIWTKPQCLAAQCPLKIP-ERGNMI--CL-HSAKAF 282
QY 138 QPKMGKRLGAALKRHSKCNCRGCLANYCEYEAKIMCSS-----ICKCIACK 188
Db 283 Q-----HQSACS---ESCEGFALVGPVVQCTASGVTAPAPVCAVQCQ 325
QY 189 NYEESPERKMLSTHYMEPGDFESSHYLSPAKFSPPKLR-KNQAFSCIS---WEVVE 244
Db 326 HLEAFSEGTMDCVHP-----LTAFYGGSCKFECQPGYRVRGLDMLRCIDSGHWSAPL 378
QY 245 ATCACLAAQGEAE---QBHCSPSL 266
Db 379 PTCEALSCPELSPVHGSMDCSPSL 403

RESULT 9
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Slahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbets, Scott
; TITLE OF INVENTION: ICAM-1/LEA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513
; FILING DATE: 19-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: (816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 541..581
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 582..617
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 701..723
; OTHER INFORMATION: /label= trans
; OTHER INFORMATION: /note= "transmembrane domain"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 724..769
; OTHER INFORMATION: /label= cyto
; OTHER INFORMATION: /note= "cytoplasmic domain"
; PUBLICATION INFORMATION:
; AUTHORS: Pigott,
; TITLES: LFA-1 Amino acid sequence (B2) (from human
; TITLES: tonsil)
; JOURNAL: The Adhesion Molecule Facts Book
; PAGES: 96-96
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
; US-08-789-078-1

-Query Match 5.3%; Score 85.5; DB 2; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;
```



```
; MOLECULE TYPE: protein
US-08-938-085A-31

Query Match          5.3%; Score 85.5; DB 4; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

QY 90 PKRITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELEFRKAIRKAC 129
Db 525 PGKLIYGQYCECDTINCERYNGQVCGPGRGLFCGKCRCHPGFEGSACQCR--TTEGC 582
QY 130 LDRNPEAFQPMKGRLGAALKRHSKGNCRSGCLKNKYCEYAKIM-----CSSI 181
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGYQLPLCQECPCPSP 621

QY 182 C-----KCIACKNYEESP 194
Db 622 CGKYISCAECLKFEKGP 638

RESULT 14
PCT-US95-04886-1
; Sequence 1, Application PC/TUS9504886
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Shahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbets, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: (816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
```

```
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 541..581
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 582..617
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 701..723
; OTHER INFORMATION: /label= trans
; OTHER INFORMATION: /note= "transmembrane domain"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 724..769
; OTHER INFORMATION: /label= cyto
; OTHER INFORMATION: /note= "cytoplasmic domain"
; PUBLICATION INFORMATION:
; AUTHORS: Pigott,
; TITLE: LFA-1 Amino acid sequence (B2) (from human
; TITLE: tonsil)
; JOURNAL: The Adhesion Molecule Facts Book
; PAGES: 96-96
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

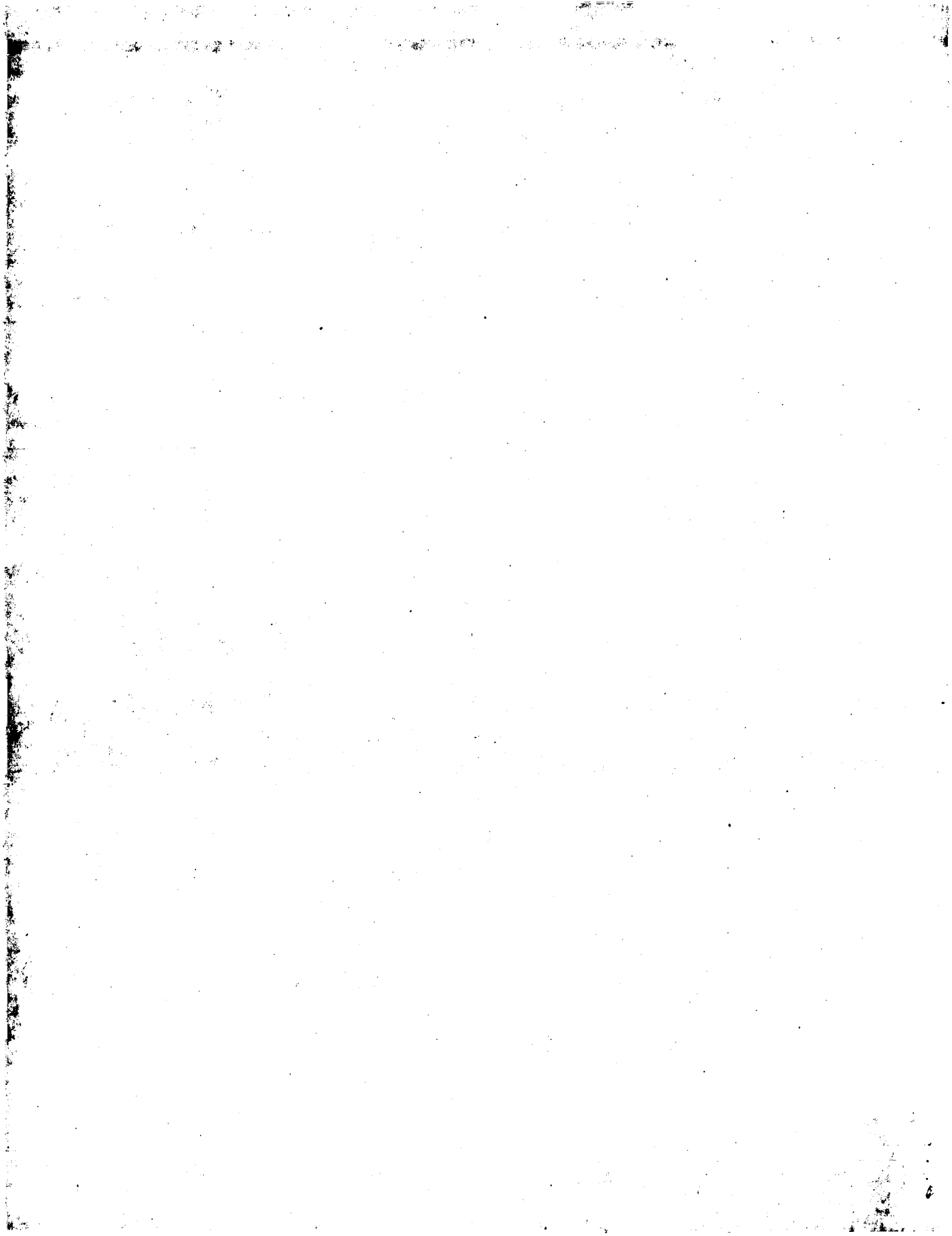
Query Match          5.3%; Score 85.5; DB 5; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

QY 90 PKRITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELEFRKAIRKAC 129
Db 525 PGKLIYGQYCECDTINCERYNGQVCGPGRGLFCGKCRCHPGFEGSACQCR--TTEGC 582
QY 130 LDRNPEAFQPMKGRLGAALKRHSKGNCRSGCLKNKYCEYAKIM-----CSSI 181
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGYQLPLCQECPCPSP 621

QY 182 C-----KCIACKNYEESP 194
Db 622 CGKYISCAECLKFEKGP 638

RESULT 15
PCT-US96-01314-45
; Sequence 45, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaout
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
```

Search completed: April 21, 2003, 11:51:09
Job time : 16.4091 secs



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	396	24.8	438	10	US-09-220-091-13	Sequence 13, Appl
2	115	7.2	1192	9	US-10-189-971-18	Sequence 18, Appl
3	115	7.2	1207	9	US-10-189-971-20	Sequence 20, Appl
4	115	7.2	1477	9	US-10-189-971-8	Sequence 8, Appl
5	115	7.2	1535	9	US-10-189-971-14	Sequence 14, Appl
6	115	7.2	1593	9	US-10-189-971-4	Sequence 4, Appl
7	107.5	6.7	1057	9	US-10-189-971-6	Sequence 6, Appl
8	107.5	6.7	1251	9	US-10-189-971-16	Sequence 16, Appl
9	107.5	6.7	1342	9	US-10-189-971-24	Sequence 24, Appl
10	107.5	6.7	1512	9	US-10-189-971-10	Sequence 10, Appl
11	107.5	6.7	1570	9	US-10-189-971-12	Sequence 12, Appl
12	107.5	6.7	1628	9	US-10-189-971-2	Sequence 2, Appl
13	100	6.3	759	9	US-10-189-971-22	Sequence 22, Appl
14	97.5	6.1	652	10	US-09-789-919-96	Sequence 96, Appl
15	96	6.0	3635	9	US-10-037-182-4	Sequence 4, Appl
16	96	6.0	3635	10	US-09-845-583-2	Sequence 2, Appl
17	93	5.8	4123	9	US-10-213-509-5	Sequence 5, Appl
18	91.5	5.7	644	10	US-09-789-919-62	Sequence 62, Appl
19	91.5	5.7	644	10	US-09-789-919-73	Sequence 73, Appl

Db 367 EAATMCMVAQAEALNVEKQVTEDEKLINMEKLVLEFRGRCLEQMI 412

RESULT 2

US-10-189-971-18
; Sequence 18, Application US/10189971
; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/315,634
; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 18

; LENGTH: 1192

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-18

Query Match 7.2%; Score 115; DB 9; Length 1192;

Best Local Similarity 21.1%; Pred. No. 0.1;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 352 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 407

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTKITSYGC-----DC-----F 102

Db 408 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 445

QY 103 SSGD-----FCNSCSCNNLRHLEFRFAKAKACLDNPEAFQPKMK-----GRL 146

Db 446 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEPGAGSCWCRCQAGV 492

QY 147 GAALKR-----HSKGCNCKR--SGCLK-----NYCECYE----- 173

Db 493 SCVRLQCPPLPCKLQVTERGSCCRGCLAHGEEHPEGSRWVPPDSACSCVCHEGVTT 552

QY 174 -AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMEPGD--FESSHYLSPAKF 222

Db 553 CARIQCISSCAQPROGPHDCPCDCGCE-----YEPGESFQPG--ADPCEV 601

QY 223 -----SGPPKLRNQAFCISWEVVEATCACLQAAGEAEQHCSPSLAEOM 270

Db 602 CICEPQEPGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----PQHCCPTCAEAL 648

RESULT 3

US-10-189-971-20

; Sequence 20, Application US/10189971

; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/315,634

; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20

; LENGTH: 1207

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-20

Query Match 7.2%; Score 115; DB 9; Length 1207;

Best Local Similarity 21.1%; Pred. No. 0.1;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 367 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 422

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTKITSYGC-----DC-----F 102

Db 423 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 460

QY 103 SSGD-----FCNSCSCNNLRHLEFRFAKAKACLDNPEAFQPKMK-----GRL 146

Db 461 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEPGAGSCWCRCQAGV 507

QY 147 GAALKR-----HSKGCNCKR--SGCLK-----NYCECYE----- 173

Db 508 SCVRLQCPPLPCKLQVTERGSCCRGCLAHGEEHPEGSRWVPPDSACSCVCHEGVTT 567

QY 174 -AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMEPGD--FESSHYLSPAKF 222

Db 568 CARIQCISSCAQPROGPHDCPCDCGCE-----YEPGESFQPG--ADPCEV 616

QY 223 -----SGPPKLRNQAFCISWEVVEATCACLQAAGEAEQHCSPSLAEOM 270

Db 617 CICEPQEPGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----PQHCCPTCAEAL 663

RESULT 4

US-10-189-971-8

; Sequence 8, Application US/10189971

; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 1477

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-8

Query Match 7.2%; Score 115; DB 9; Length 1477;

Best Local Similarity 21.1%; Pred. No. 0.13;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 637 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 692

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTKITSYGC-----DC-----F 102

Db 693 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCPTCGCRYHGVTT 730
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 731 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEGPAGSCWCRCAQGV 777
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCEC----- 173
Db 778 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSSACSCVCHGVTT 837
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 838 CARIQCISSCAQPRGPHDCCPQCSDEHEGRK-----YEPGESFQPG--ADPCEV 886
Qy 223 -----SGPPKLRKRNQAFSCISWEVVEATCACLQAQGEAEQHCSPSLAEOM 270
Db 887 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 933

RESULT 5
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucleot
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-14

Query Match 7.28; Score 115; DB 9; Length 1535;
Best Local Similarity 21.18; Pred. No. 0.14;
Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

Qy 2 VICOLKGAQML-----CIDMCGARELKALHLLPOYDQSSFPQSELPKPMTTLVGR 53
Db 695 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPCNLTCLGG 750
Qy 54 LLPVPAKLNLTQVNDGALPSAVNGAAPSGPALQGPPIKTLGYC-----DC-----F 102
Db 751 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCPTCGCRYHGVTT 788
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 789 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEGPAGSCWCRCAQGV 835
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCEC----- 173
Db 836 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSSACSCVCHGVTT 895
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 896 CARIQCISSCAQPRGPHDCCPQCSDEHEGRK-----YEPGESFQPG--ADPCEV 944
Qy 223 -----SGPPKLRKRNQAFSCISWEVVEATCACLQAQGEAEQHCSPSLAEOM 270
Db 945 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 991

RESULT 6
US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Query Match 7.28; Score 115; DB 9; Length 1593;
Best Local Similarity 21.18; Pred. No. 0.14;
Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

Qy 2 VICOLKGAQML-----CIDMCGARELKALHLLPOYDQSSFPQSELPKPMTTLVGR 53
Db 753 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPCNLTCLGG 808
Qy 54 LLPVPAKLNLTQVNDGALPSAVNGAAPSGPALQGPPIKTLGYC-----DC-----F 102
Db 809 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCPTCGCRYHGVTT 846
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 847 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEGPAGSCWCRCAQGV 893
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCEC----- 173
Db 894 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSSACSCVCHGVTT 953
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 954 CARIQCISSCAQPRGPHDCCPQCSDEHEGRK-----YEPGESFQPG--ADPCEV 1002
Qy 223 -----SGPPKLRKRNQAFSCISWEVVEATCACLQAQGEAEQHCSPSLAEOM 270
Db 1003 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 1049

RESULT 7
US-10-189-971-6
; Sequence 6, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0

```
; SEQ ID NO 6
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-6

Query Match
Best Local Similarity 20.0%; Score 107.5; DB 9; Length 1057;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
DB 182 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNQEFFDPREPCNLCTCLGG 237
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
DB 238 FVTGRRRPPCGSHPLPSGHCCPTCGCRVHGVTASGETLPDLPDPTCSLCTCQEG 297
QY 82 -----PSGPALQGGPKITLSCYCDGFCSSGDFCNSCNCNNLRHELEKFAIKACLDNRN 134
DB 298 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSGQREHQ-----DG 337
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
DB 338 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 397
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
DB 398 RWVPPDSACSSVCHEGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 450
QY 204 HYMEPGD-FESSHYLSPAKF-----SGPPKLRKQAFSCISWEVVEATCACLIAQGE 255
DB 451 --YEPGESFQPG--ADPCEVCICEPOEGPPSLRCHRR--QCPS--LVGCPPSQLLPPG- 501
QY 256 EAEQEHCSPLAQQM 270
DB 502 ---PQHCCPTCAEAL 513

RESULT 8
US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleoti
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

Query Match
Best Local Similarity 20.0%; Score 107.5; DB 9; Length 1251;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
DB 376 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNQEFFDPREPCNLCTCLGG 431
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
```

```
DB 432 FVTGRRRPPCGSHPLPSGHCCPTCGCRVHGVTASGETLPDLPDPTCSLCTCQEG 491
QY 82 -----PSGPALQGGPKITLSCYCDGFCSSGDFCNSCNCNNLRHELEKFAIKACLDNRN 134
DB 492 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSGQREHQ-----DG 531
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
DB 532 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 591
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
DB 592 RWVPPDSACSSVCHEGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 644
QY 204 HYMEPGD-FESSHYLSPAKF-----SGPPKLRKQAFSCISWEVVEATCACLIAQGE 255
DB 645 --YEPGESFQPG--ADPCEVCICEPOEGPPSLRCHRR--QCPS--LVGCPPSQLLPPG- 695
QY 256 EAEQEHCSPLAQQM 270
DB 696 ---PQHCCPTCAEAL 707

RESULT 9
US-10-189-971-24
; Sequence 24, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1342
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-24

Query Match
Best Local Similarity 20.0%; Score 107.5; DB 9; Length 1342;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
DB 467 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNQEFFDPREPCNLCTCLGG 522
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
DB 523 FVTGRRRPPCGSHPLPSGHCCPTCGCRVHGVTASGETLPDLPDPTCSLCTCQEG 582
QY 82 -----PSGPALQGGPKITLSCYCDGFCSSGDFCNSCNCNNLRHELEKFAIKACLDNRN 134
DB 583 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSGQREHQ-----DG 622
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
DB 623 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 682
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
DB 683 RWVPPDSACSSVCHEGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 735
```


QY 204 HYPEGD-FESSHYLSPAKF-----SGPPKLKRNRQAFSCISWEVVEATCACLIAQGE 255
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 736 --YEPGESFQPG--ADPCEVCICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 786
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 256 EAEQEHCSPLABQM 270
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 787 ---PQHCCPTCAEAL 798
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 10

US-10-189-971-12
; Sequence 10, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1512
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-10

Query Match 6.7%; Score 107.5; DB 9; Length 1512;
Best Local Similarity 20.0%; Pred. No. 0.61;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPQYDDOSSFPQSELPKPMTTLVGR 53
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 637 VSCEPKACAPALCPPARGDCPCDCGCE----YLGESYLSNQEFFDPDPREPCNLCTCLGG 692
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 54 LLP-----VPA-----KLNLTQVNDGALPSAVNGAAF----- 81
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 693 FVTCGRRCPEPGCSHPLIPSGHCCPTCQCGRYHGVTTASGETLPDLPDPTCSLCTCQEG 752
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 82 -----PSGPALQGPPIKTLGYCDCFSSGDFCNSCNCNNLRHELEKAKIACLDNRN 134
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 753 SMRCQKKCAPALCPHPS---PGPCFC----PVCHSCLSQGREHQ-----DG 792
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR--SGCLK----- 166
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 793 EEFEGPAGSCWCRCQAGVSVRLQCPPLCKLQVTERGSCPCRCGCLAHGEHPEGS 852
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 853 RWVPPDSACSSVCHEGVVTCARIOCISSCAOPRGPHDCCPCQCDCEHGRK----- 905
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 204 HYPEGD-FESSHYLSPAKF-----SGPPKLKRNRQAFSCISWEVVEATCACLIAQGE 255
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 906 --YEPGESFQPG--ADPCEVCICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 956
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 256 EAEQEHCSPLABQM 270
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 957 ---PQHCCPTCAEAL 968
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 11

US-10-189-971-12
; Sequence 12, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade

; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1570
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-12

Query Match 6.7%; Score 107.5; DB 9; Length 1570;
Best Local Similarity 20.0%; Pred. No. 0.63;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPQYDDOSSFPQSELPKPMTTLVGR 53
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 695 VSCEPKACAPALCPPARGDCPCDCGCE----YLGESYLSNQEFFDPDPREPCNLCTCLGG 750
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 54 LLP-----VPA-----KLNLTQVNDGALPSAVNGAAF----- 81
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 751 FVTCGRRCPEPGCSHPLIPSGHCCPTCQCGRYHGVTTASGETLPDLPDPTCSLCTCQEG 810
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 82 -----PSGPALQGPPIKTLGYCDCFSSGDFCNSCNCNNLRHELEKAKIACLDNRN 134
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 811 SMRCQKKCAPALCPHPS---PGPCFC----PVCHSCLSQGREHQ-----DG 850
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR--SGCLK----- 166
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 851 EEFEGPAGSCWCRCQAGVSVRLQCPPLCKLQVTERGSCPCRCGCLAHGEHPEGS 910
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 911 RWVPPDSACSSVCHEGVVTCARIOCISSCAOPRGPHDCCPCQCDCEHGRK----- 963
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 204 HYPEGD-FESSHYLSPAKF-----SGPPKLKRNRQAFSCISWEVVEATCACLIAQGE 255
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 964 --YEPGESFQPG--ADPCEVCICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 1014
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 256 EAEQEHCSPLABQM 270
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1015 ---PQHCCPTCAEAL 1026
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

RESULT 12

US-10-189-971-2
; Sequence 2, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1628

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-2

Query Match
Best Local Similarity 6.78; Score 107.5; DB 9; Length 1628;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

Qy 2 VICOLKGAOML-----CIDNCGARELKHLLPOYDDQSSFPQSELPKPMWTLVGR 53
Db 753 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPCNLTCLGG 808
Qy 54 LLP-----VPA-----KLNLIQVDNGALPSAVNGAAF-----81
Db 809 FVTCGRRRCEPPGSHPLIPSGHCHCPTCQCGRYHGVTTASGETLPDPLDPTCSLCTCQEG 868
Qy 82 -----PSGALOGPPKITLSGYCDGFCSSGDFNCSCNNLRHELFRFAKACLDNRN 134
Db 869 SMRCQKRCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 908
Qy 135 EAFQPKMGK-----GRIGAALKR-----HSKGCNCRK-SGCLK-----166
Db 909 BEFEGPAGCWCRCQAGQVSCVRLQCPKLVQTERGSCCPRCGRCLAHGEHPEGS 968
Qy 167 -----NYCECYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 969 RWPVPDSACSCVCHEGVVTCARIQCISSCAQPRGPHDCCPQSCDCEHEGRK-----1021
Qy 204 HYMEPGD-FESHYLSPAKF-----SGPPKLRKNQAFSCISWEVVYATCACLQAQE 255
Db 1022 --YEPGESFQPG--ADPCEVCICEPQEPGPPSLRCHRR--QCPS--LVGCPSPQLLPPG- 1072
Qy 256 EAEQEHCSPSLAEQM 270
Db 1073 ---FOHCCPTCAEAL 1084

RESULT 13
US-10-189-971-22
; Sequence 22, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 759
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-22

Query Match
Best Local Similarity 6.38; Score 100; DB 9; Length 759;
Matches 56; Conservative 24; Mismatches 70; Indels 98; Gaps 16;

Qy 82 PSGALOGPPKITLSGYCDGFCSSGDFNCSCNNLRHELFRFAKACLDNRNPEAFQPKM 141
Db 7 PCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DGEFEFGPA 46
Qy 142 GK-----GRIGAALKR-----HSKGCNCRK-SGCLK-----166
Db 47 GSCWCRCQAGQVSCVRLQCPKLVQTERGSCCPRCGRCLAHGEHPEGSRWVPPDS 106

; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-22

Query Match
Best Local Similarity 6.78; Score 107.5; DB 9; Length 1628;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

Qy 2 VICOLKGAOML-----CIDNCGARELKHLLPOYDDQSSFPQSELPKPMWTLVGR 53
Db 753 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPCNLTCLGG 808
Qy 54 LLP-----VPA-----KLNLIQVDNGALPSAVNGAAF-----81
Db 809 FVTCGRRRCEPPGSHPLIPSGHCHCPTCQCGRYHGVTTASGETLPDPLDPTCSLCTCQEG 868
Qy 82 -----PSGALOGPPKITLSGYCDGFCSSGDFNCSCNNLRHELFRFAKACLDNRN 134
Db 869 SMRCQKRCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 908
Qy 135 EAFQPKMGK-----GRIGAALKR-----HSKGCNCRK-SGCLK-----166
Db 909 BEFEGPAGCWCRCQAGQVSCVRLQCPKLVQTERGSCCPRCGRCLAHGEHPEGS 968
Qy 167 -----NYCECYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 969 RWPVPDSACSCVCHEGVVTCARIQCISSCAQPRGPHDCCPQSCDCEHEGRK-----1021
Qy 204 HYMEPGD-FESHYLSPAKF-----SGPPKLRKNQAFSCISWEVVYATCACLQAQE 255
Db 1022 --YEPGESFQPG--ADPCEVCICEPQEPGPPSLRCHRR--QCPS--LVGCPSPQLLPPG- 1072
Qy 256 EAEQEHCSPSLAEQM 270
Db 1073 ---FOHCCPTCAEAL 1084

RESULT 14
US-09-789-919-96
; Sequence 96, Application US/09789919
; Patent No. US2002006485A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; TITLE OF INVENTION: CELLS AND USES THEREOF
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 96
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-96

Query Match
Best Local Similarity 6.18; Score 97.5; DB 10; Length 652;
Matches 31; Conservative 12; Mismatches 49; Indels 19; Gaps 4;

Qy 96 SGYC--DCFSSGDFCNSCSCNNLRHELFRFAKACLDNRNPEAFQPKMGKRLGAALKRH 153
Db 268 NGGCHQDCFEFGDGSFLCGCRPGRLLD---DLVTCASRNPCSSPCRG-----313
Qy 154 SKGCNCRKRSGLKNY-CEYEAKIMCSICKYACIMCKNYEESPERKMLMSTP 203
Db 314 --GATCVLGHGKNYKRCRCPOGYQLDSSQLDCVDVDECQDSPCAQECVNTP 362

RESULT 15
US-10-037-182-4
; Sequence 4, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thybrell, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-182-4

Query Match
Best Local Similarity 6.08; Score 96; DB 9; Length 3635;
Matches 18.08; Pred. No. 18;
```

Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

```
QY 14 CTDNCGARELKALHLLPQYDDOSSFPQSELPKPMTLYGRLLLPVPAKLNLIITQVDNGALP 73
Db 117 CLERFGPRTLER---ITODDDVI-----CTEYSRIVPL-ENGEIVVSVLNGR-P 161
QY 74 SAVNGAAFP-----SGPALQGP-----PKITLSGY 98
Db 162 GALNFSYSPLLRDFTKATNIRLRLFTNTLLGHLMGKALRDPVTYRRYYYSIKDISIGR 221
QY 99 CDCFSSGDFC-----NSC-----SC 113
Db 222 CYCHGHADYCDAKDPLDPFLQACQHNTCGGSCDRCCPGFNQOPWKPATTDSANECQSC 281
QY 114 NNLRH-----ELERFKAIKACLDRNPEAFQPKMGKRLGAAL---RHSKGCNCKRS 162
Db 282 NCHGHAYDCYDPEVDR-----RNASQNDNVYOG--GGVCLDCQHHTGTINCER- 329
QY 163 GCLKNY-----CECYEAKI---MCSSI-----CK-----CIAC-KNY 190
Db 330 -CLPGFFRAPDQPLDSPHVCRPCDC-ESDFTDGTCTEDLTGRCYCRPNFTGELCAACABGY 387
QY 191 EESPERKMLMSTPHYMEPGDPESHYLSPAKFSGPPKLRKNRQAFSCISWEVVEATCACL 250
Db 388 TDFPHCYPLPSFPHN-----DTREQVLPA-----GQIVNCDCNAA 422
QY 251 LAQGEAEQE-----HC-----SPSLAEQMILEEFG 277
Db 423 GTQGNACRKDPLGRGVCKPNFRGAHCELCAFGFHGSPCHPCQCSSPGVANSLCDPESGQ 482
QY 278 CL 279
Db 483 CM 484
```

Search completed: April 21, 2003, 11:51:47
Job time : 19.4024 secs

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 210.79 Seconds
(without alignments)
492.596 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLHIEFRKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
- 26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	100.0	295	21	US-09-743-237-4
2	1599	100.0	295	21	US-09-743-237-4
3	1251	78.2	299	21	US-09-743-237-5
4	505.5	31.6	223	1	PCT-US01-08656-10693
5	435	27.2	403	26	US-60-161-932-1700
6	435	27.2	890	26	US-60-167-217-12948
7	435	27.2	950	20	US-09-614-150-12897

Query Match	100.0%	Score	1599;	DB	21;	Length	295;
Best Local Similarity	100.0%	Pred. No.	1.3e-143;				
Matches	295;	Conservative	0;	Mismatches	0;	Indels	0;
QY	1	MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK	60				
DB	1	MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK	60				

RESULT 1

US-09-743-237-4
; Sequence 4, Application US/097433237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-743-237-4

ALIGNMENTS

8	435	27.2	950	26	US-60-191-637-12935	Sequence 12935, A
9	396	24.8	438	14	US-09-087-136-13	Sequence 13, Appl
10	396	24.8	438	14	US-09-087-136-13	Sequence 13, Appl
11	396	24.8	438	16	US-09-220-091-13	Sequence 13, Appl
12	363	22.7	550	21	US-09-708-427-33366	Sequence 33366, A
13	363	22.7	571	21	US-09-708-427-33365	Sequence 33365, A
14	362	22.6	280	1	PCT-US01-08631-48317	Sequence 48317, A
15	357	22.3	524	21	US-09-708-427-33367	Sequence 33367, A
16	357	22.3	603	19	US-09-573-655A-2154	Sequence 2154, Ap
17	357	22.3	603	19	US-09-573-655A-2152	Sequence 2152, Ap
18	357	22.3	603	21	US-09-708-427-27294	Sequence 27294, A
19	352.5	22.0	534	21	US-09-708-427-27295	Sequence 27295, A
20	351.5	22.0	615	21	US-09-733-089-22002	Sequence 22002, A
21	351.5	22.0	615	21	US-09-733-089-22183	Sequence 22183, A
22	351.5	22.0	615	21	US-09-733-089-22184	Sequence 22184, A
23	351.5	22.0	615	22	US-09-816-660-22002	Sequence 22002, A
24	351.5	22.0	615	22	US-09-816-660-22183	Sequence 22183, A
25	351.5	22.0	615	22	US-09-816-660-22184	Sequence 22184, A
26	333.5	20.9	518	21	US-09-708-427-27296	Sequence 27296, A
27	330	20.6	241	26	US-60-324-109-27382	Sequence 27382, A
28	322	20.1	147	1	PCT-US01-01329-1409	Sequence 1409, Ap
29	322	20.1	147	1	PCT-US01-01339-3988	Sequence 3988, Ap
30	322	20.1	147	21	US-09-764-891-3988	Sequence 3988, Ap
31	322	20.1	147	24	US-10-080-090-1409	Sequence 1409, Ap
32	276.5	17.3	207	21	US-09-733-089-22013	Sequence 22013, A
33	276.5	17.3	207	22	US-09-816-660-22013	Sequence 22013, A
34	276	17.3	168	21	US-09-733-089-21980	Sequence 21980, A
35	276	17.3	168	22	US-09-816-660-21980	Sequence 21980, A
36	248	15.5	53	1	PCT-US00-26524B-6922	Sequence 6922, Ap
37	244	15.3	349	18	US-09-428-944-1881	Sequence 1881, Ap
38	234.5	14.7	386	26	US-60-324-109-16915	Sequence 16915, A
39	234.5	14.7	556	26	US-60-324-109-32886	Sequence 32886, A
40	232	14.5	497	21	US-09-708-427-25631	Sequence 25631, A
41	232	14.5	542	21	US-09-708-427-25630	Sequence 25630, A
42	232	14.5	658	21	US-09-708-427-25629	Sequence 25629, A
43	230.5	14.4	770	26	US-60-324-109-17053	Sequence 17053, A
44	230.5	14.4	770	26	US-60-324-109-17063	Sequence 17063, A
45	219.5	13.7	402	26	US-60-324-109-31233	Sequence 31233, A


```
Db      58 ASRPKPCNCTSLCLUKLYCDGFANGFCNNCNTNYNLEHENERQKAICACLDNRPE 117
QY     136 AFOPKMKGRLGAALKRHSGKNCKRSGCLKKYCEYEAKIMCSSICKCAIACKNYESPE 195
       ||::|||: | : ||||||||||||||||||||||||||||| |||::|||
Db     118 AFRPKIGKEGSDRRHSKGCNKCRSGCLKKYCEYEAKIMCSSICKCIGCNFEESPE 177
QY     196 RKWLM 200
       || ||
Db     178 RKTLM 182

RESULT   5
US-60-161-932-1700
; Sequence 1700, Application US/60161932
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; ; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences
; ; FILE REFERENCE: CL000122
; ; CURRENT APPLICATION NUMBER: US/60/161,932
; ; CURRENT FILING DATE: 1999-10-28
; ; NUMBER OF SEQ ID NOS: 2626
; ; SOFTWARE: FastSeq for Windows Version 4.0
; ; SEQ ID NO 1700
; ; LENGTH: 403
; ; TYPE: PRT
; ; ORGANISM: Drosophila
US-60-161-932-1700

Query Match          27.2%; Score 435; DB 26; Length 403;
Best Local Similarity 38.4%; Pred.No. 2.6e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

QY    31 QYDQSFPQSELPRPMTTLVGRLLPVPAKLNLITQVDNGALPSAVNGAAFGPGALQGP 90
      + : + : ||| ||| ||| : + : ||| : + : |||
Db   134 QLPTQEOSTPIKVEPKLPTLPVGKANVPAP--PLFVLKPPTAAAAGAVDPGLGWTSSR 191
QY    91 PKTILSG-----YCOCFSSEDFNCSCS----NNLRHELERFKAIKACLDRNPFAQPK 140
      |||||::||| |||||::||| |||||::||| |||||::||| |||||
Db   192 KHCNCSKSQCCLKLYDCFANGFECODCTCKDCFNNLDYEVEVERAIRSCLDNRPSAFKPK 251
QY   141 MGKRGRLGAALKRHSGKNCKRSGCLKKYCEYEAKIMCSSICKCAIACKNYESPERKMLM 200
      : + : ||| ||||| ||||| ||||| ||||| ||||| ::||| :
Db   252 ITAPNSGDWRLL-HNKCNCNKRSGCLKKYCEYEAKIPCSSICKVCGRNMEDRPD----- 305
QY   201 STPHYPEPDGFESSHYLSPAKSPGGPPKLRKNRQ-----AFSCISWEWEATCACILAQQE 255
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db   306 -----VMDSDLGLMGVEGKKDK-AKKQLNENRANIYFTDDVIETAMCMISRIV 356
QY   256 EAEQHCHSPSLARQMILEEFRCCLSOIL 283
      ||::: | : ||| | : |||
Db   357 MHEKQNVAVEDMEREYMEEENGESLTQII 384

RESULT   6
US-60-167-217-12948
; Sequence 12948, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; ; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; ; FILE REFERENCE: CL000152
; ; CURRENT APPLICATION NUMBER: US/60/167,217
; ; CURRENT FILING DATE: 1999-11-24
; ; NUMBER OF SEQ ID NOS: 23195
; ; SOFTWARE: FastSeq for Windows Version 4.0
; ; SEQ ID NO 12948
; ; LENGTH: 890
; ; TYPE: PRT
; ; ORGANISM: Drosophila
```

```

US-60-167-217-12948

Query Match          27.2%; Score 435; DB 26; Length 890;
Best Local Similarity 38.4%; Pred. No. 7.8e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps

QY 31 QYDQSSFPQSELPKPMTTLLVGRLLPPVPAKLNLIITVDNGALPSAVNGAAPPSPGALQGP 90
DB 621 QLPTQSTPIKVPKPLPTLPPGVKANVPK--PLFEVLKPPATAAAGAVDPLGGMTSRR 678

QY 91 PKITLSG-----YDCFSGDFCNSCSC-----NNLRHELERFKATKACLDNRNPEAFQPK 140
DB 679 KHCNCSKSQCLKLYCDCFANGFECQDCTCKDCFNLDYEVEVERAIRATRSCLDRNPISAFKPK 738

QY 141 MGKRGRLGAALKRISKGCNCRSLCKNKCCEYBAKIMCWSICCKIACKNVEESPERKMLM 200
DB 739 ITAPNSGDMRL-HNKGCNCRSLCKNKCCEYBAKIPCCSICKVCVCRNMEDRPD----- 792

QY 201 STPHMEYPGDFESSHLYSPAKFSGPPKLRNRO-----AFSCISWEVEATCACLLAOG 255
DB 793 -----VDMDSLGLMGVGGOKDK-ANKQLNENRANIYFTDDVIEATIMCMISRI 843

QY 256 EAEQHCSPSLAEOMILEEFGRLCSQIL 283
DB 844 MHEKQNAVDEMEREVMEENGESITQII 871

RESULT 7
US-09-614-150-12897
: Sequence 12897, Application US/09614150
: GENERAL INFORMATION:
: 'APPLICANT: Venter, J. Craig
: APPLICANT: et al.
: TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
: TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
: TITLE OF INVENTION: DROSOPHILA GENES.
: FILE REFERENCE: CL000728
: CURRENT APPLICATION NUMBER: US/09/614,150
: CURRENT FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/157,832
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: 60/160,191
: PRIOR FILING DATE: 1999-10-19
: PRIOR APPLICATION NUMBER: 60/161,932
: PRIOR FILING DATE: 1999-10-28
: PRIOR APPLICATION NUMBER: 60/164,769
: PRIOR FILING DATE: 1999-11-12
: PRIOR APPLICATION NUMBER: 60/173,383
: PRIOR FILING DATE: 1999-12-28
: PRIOR APPLICATION NUMBER: 60/175,693
: PRIOR FILING DATE: 2000-01-12
: PRIOR APPLICATION NUMBER: 60/184,831
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: 60/191,637
: PRIOR FILING DATE: 2000-03-23
: NUMBER OF SEQ ID NOS: 43008
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 12897
: LENGTH: 950
: TYPE: PRT
: ORGANISM: DROSOPHILA
US-09-614-150-12897

Query Match          27.2%; Score 435; DB 20; Length 950;
Best Local Similarity 38.4%; Pred. No. 8.5e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps

QY 31 QYDQSSFPQSELPKPMTTLLVGRLLPPVPAKLNLIITVDNGALPSAVNGAAPPSPGALQGP 90
DB 681 QLPTQSTPIKVPKPLPTLPPGVKANVPK--PLFEVLKPPATAAAGAVDPLGGMTSRR 938

```

	Query Match	27.2%	Score 435	DB 20	Length 950
Best Local	Similarity 38.4%		Pred. NO. 8.5e-32		
Matches 103	Conservative 40		Mismatches 32		
			Indels 32		
Qy	31 QYDDSSFFPSELPKMTTLVGRLLPVPKLNILITVDNGALPSAVNGAFPSGPALOGP 90				
Db	681 QLPTQSTPIKVEPKLPTLPVGKANVPK--PLFVLKPPATAAAGAVDPLGGMTSR 738				

Qy	36	SSFFQSSELPRPMTTLVGRLL-----PVPAKLMLITQVD-NGALPSAVNGAAF	81
Dd	25	SMAAAISPISIVTVRPIITSQAAPTVAIPPPSQGIIILHVTRHPRESPNSMPR	84
Qy	82	PSGPALOGPKP-----ITLSGYCDCFSSGDFCNCSGC---NNLRHEUERKATK	127
Dd	85	PAGETRGTPQKKQCCKHSCRCLKLYCECFASGTYDGCNCVCNNVNEVEPARQAVE	144
Qy	128	ACLDRNPFAOPKMGKGRLGAAK-----LRHSKCNCCKRSGLCNKYCECYEAKI	176
Dd	145	STLERNPNAFRPKTAASPFGHGRRNRREBVGWVMLARHNKGCHCKKSGCLKYCECFQANI	204
Qy	177	MCSSTICKCIACKNYEESPERRKLMLSTPH----YME-----PGDESSHVLSPANKFSGP	225
Dd	205	LCSNCKLDCKNFEGSEVRQSLFHGEHSHNLAY/LQHANAATGATCGSGFAS----APP	260
Qy	226	PKLRKNRAQF	235
Dd	261	PKRKKGQEIF	270

RESULT 13
US-09-708-427-33365
Sequence 33365, Application US/09708427
GENERAL INFORMATION:
APPLICANT: N. ALEXANDROV et al.
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING
TITLE OF INVENTION: THEREBY
FILE REFERENCE: 2750-1243P
CURRENT APPLICATION NUMBER: US/09/708,427
CURRENT FILING DATE: 2000-11-09
NUMBER OF SEQ ID NOS: 85364
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33365
LENGTH: 571
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..571
OTHER INFORMATION: xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: 1..571
OTHER INFORMATION: Ceres Seq. ID 1836304
US-09-708-427-33365

[illegible]

RESULT 14

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: Thereby
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2192
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2192

Query Match 22.3%; Score 357; DB 5; Length 603;
Best Local Similarity 35.2%; Pred. No. 8.6e-24;
Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;
QY 36 SFPQSELPRMTLVGRLLPVPAKLMLITQVDNGALPSAVNGA----AFPSGPAQGP- 90
DB 60 SSVPESTIRP-GMTAIGQVTVQRTPLPMATMSNPSPQSOIVNAPIRHPIPESPARGPR 118
QY 91 PKI-----TISGYCDGCFSSGDFCNSCSC-----NLRHLELERPKA 127
DB 119 PNVEGRDGTQKKQCKNSRCLKLYCECFASGTGDCGNCVCFNNVDNEPARREAVE 178
QY 128 ACLDRNPEAFQPMKGK----GR-----LGAALKL--RHSGKCNCKRSGCLKNYCEYEAKI 176
DB 179 ATLERNPFAFRPKIASPHGGRDKREDIGEVLLGRHNKCHCKKSGCLKLYCECFQANI 238
QY 177 MCSICKIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYL-SPAKFS 223
DB 239 LCSENCKLCKNPEGSEERQALFHGSHNMAVYLAQAANAATGAVGSSGFAPSPA--- 295
QY 224 GPKLRLNRQ 233
DB 296 --PKRRKGQE 303

RESULT 3
US-10-155-881-9626
; Sequence 9626, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: TRANSCRIPTION IN PLANTS
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9626
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9626

Query Match 20.9%; Score 333.5; DB 6; Length 282;
Best Local Similarity 37.4%; Pred. No. 4.3e-22;
Matches 73; Conservative 28; Mismatches 45; Indels 49; Gaps 7;
QY 46 PMTTLVGRLLPVPAKLMLITQVDNGALPSAVNGAAPPSPGALOG-----PPK---- 92
DB 21 PLAVSVPLRLPVPPQ----PVQRPVAVPLK-----PESPKLARLYEGKDSPTTKKCC 71
QY 93 -----TISGYCDGCFSSGDFCNSCSC-----NLRHLELERPKAACLDRNPEAFQPMKGK 143
DB 72 NCRNSRCLKLYCECFASGTGDCGNCVCFNNVDNEPARREAVEAILDRNPDFAFRKIGS 131

QY 144 GRL-----GAAKLRRHSGKCNCKRSGCLKNYCEYEAKIMCSSICKIACKN 189
DB 132 SPLMHRNNEVPDPLVPG-----KHNGCHCKKSGCLKLYCECFQANILCSENCKMCDKN 187
QY 190 YEESEPERKMLMSTPH 204
DB 188 FEGSEERRNLFQGDH 202
RESULT 4
US-10-155-881-9875
; Sequence 9875, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: TRANSCRIPTION IN PLANTS
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9875
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9875

Query Match 19.6%; Score 314; DB 6; Length 510;
Best Local Similarity 35.6%; Pred. No. 4.9e-20;
Matches 72; Conservative 26; Mismatches 48; Indels 56; Gaps 7;
QY 82 PSGPALQ-GPPKITLSG-----YCDGCFSSGDFCNSCS 112
DB 52 PSMPLRPPWPEITLSMKPVVEMKSGTTPAKKKKHCNCKNSQCLKLYCECFAGDGCDCGN 111
QY 113 C----NLRHLELERFKAIRACLDRNPEAFQPMKGKGL-----GA--AKLRHSGKCN 158
DB 112 CKOCGTVENEGKQAEINNTKLRNPNAPFKPIENGPIPSVWKDAGALSHPKHNGCH 171
QY 159 CKRSGCLKNYCEYEAKIMCSSICKIACKNYEESPE-----RKMLMSTPHYM 206
DB 172 CKKSGCLKLYCECFQANILCCKNCKMCDKNYEGSEBELRSTTQGDNSCDRNNIQAANVA 231
QY 207 EPGDFESSHY-LSPAKFSGPPK 227
DB 232 LPGAIGSSGYRFSFVRKRKPE 253

RESULT 5
US-10-106-698-6932
; Sequence 6932, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
\
]
^
_
`
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
{
|
}
~
_

```

; NUMBER OF SEQ ID NO
: SOFTWARE: FastSEO f

```

; SEQ ID NO 1127
; LENGTH: 120

; TYPE: PRT

; ORGANISM: Lolium perenne

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (1)...(120)

; OTHER INFORMATION: Xaa = Any Amino Acid

US-10-138-145-1127

Query Match 9.9%; Score 159; DB 6; Length 120;

Best Local Similarity 38.7%; Pred. No. 5.9e-07;

Matches 41; Conservative 13; Mismatches 34; Indels 18; Gaps 7;

QY 98 YCOCFSSGDFCNS-CSCN---NLRLHELEFKAIAKACLDNRPEAFQPKMGKRLGA-----148

Db 16 YCEC-AAGVYCEPCSCGCLNKPHEIIVLSTRKQIEFRNPLA-APKVRILSDAAQETQE 73

QY 149 -----AKLRHSGKNCNRSGCLNKYCEYAKIMCISICKIACKN 189

Db 74 DPNTPASARHKRCNCKKSKLK-YCECYQGGVGLTNCRC-ECKN 117

RESULT 10

US-10-155-881-9182

; Sequence 9182, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 9182

; LENGTH: 200

; TYPE: PRT

; ORGANISM: Zea mays

US-10-155-881-9182

Query Match 7.3%; Score 117.5; DB 6; Length 200;

Best Local Similarity 24.7%; Pred. No. 0.0056;

Matches 39; Conservative 15; Mismatches 57; Indels 47; Gaps 4;

QY 39 PQSELKPKMTTLVG-----RLLPVP--AKNLITQVDNGALPSAVNGAAPPSPALQGGPP 91

Db 26 PQAAVPMDADYGAGATNPQWPEPDHLSPPPPPPASAPTPAAEAATAAASAVSPP 85

QY 92 KITLGYDCDFSSGDFCNSCSNNLRHELEFKAIAKACLDNRPEAFQPKMGKRLGAAL 151

Db 86 -----APAEANGNSDKRRRRADGDG---107

QY 152 RHSGKNCNRSGCLNKYCEYAKIMCISICKIACKN 189

Db 108 --CKACSKRCKRLKLYCYVYASGSHCTELCGCEPCHN 143

RESULT 11

PCT-US02-01339-6

; Sequence 6, Application PC/TUS0201339

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: LEE, Ernestine A.

; APPLICANT: WALIA, Narinder K.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: TANG, Y. Tom

; APPLICANT: YUE, Henry
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra
; APPLICANT: WARREN, Bridget A.
; APPLICANT: YAO, Monique G.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: LU, Yan
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Damiel B.
; APPLICANT: GRAUL, Richard
; APPLICANT: LU, Dyung Alina M.
; TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
; FILE REFERENCE: PI-0346 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/01339
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/262,838; 60/265,927; 60/271,196; 60/274,549; 60/334,1
; PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 1774
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7722591CDI
PCT-US02-01339-6

Query Match 7.2%; Score 115; DB 1; Length 1774;

Best Local Similarity 23.1%; Pred. No. 0.13;

Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGARELKALHLLPQYDDQSSPQSELKPKMTTLVGRLLPVPKAKNLITQVDNGALPSAV 76

Db 748 NCNA-----WLLPDLTRSAS-----VGPMEESVAHAVAAGVSRLYISGGFGVALGRL 796

QY 77 NGAAPFSGPA--LOGPPKITLSGYCD-CFS---SGD-----FCNSCSNNL---RHELER 122

Db 797 LALTLPDPDCRLSSPEACNQSGACTWCHGACLSGDOAHLGCGGSPSPMPSPECRR 856

QY 123 FKAIAKACLDNRPEAFQPKMGKRLGAALKRHSKGC-NCKRSGL-----KNYC-----169

Db 857 LRTCSCLARHPRTLQPGDGE-----ASTPRCKWCTNCPGACIGRSGCTSENDCRINQ 911

QY 170 -----ECYEAKI-----MCSSICKIACKNYEESPE-RKMLMSTPHY-----205

Db 912 REVFWAGNCGSEAAACGAACDEQCTREGCKMWTRQFKRTGTRRLSVQPTDTCFSSHLL 971

QY 206 -MEPGDFESSHYLSPAKFSGPPKLRKNROAFSCISWEVVEATCACLLAOGEEAEQEHCS 264

Db 972 NYSPMPVSS---PPLPCTPTCHLLPN-----CTS-----CLDSKGGGQHCVM 1014

QY 265 SLAEQMILEEFGRCLS 280

Db 1015 SSSLQ-----QCLS 1023

RESULT 12

PCT-US02-10812-17

; Sequence 17, Application PC/TUS0210812

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: GANDHI, Ameena R.

; APPLICANT: DELEGEANE, Angelo M.

```

Query Match      5.9%; Score 94.5; DB 1; Length 538;
Best Local Similarity 20.5%; Pred. No. 2.1;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNGARELKALHLLFOYDDQSSFOSELPKPMWTLVGRLLPVPKLNLTQVNDGALP 73
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 76 CYSNCSPKFKQLML-----FPPNLPYKEIT-----LEAFAVITQM----- 112
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 SAVNGAAPSPGALQPPKKITLSGYCOCFSSGFCN-----SCSCNNLRHEL 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 113 -----LALSIGSYDDPKK-----CQSESTC1MNPVQSNVGKTFSSCSLRSFQNF 161
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 ERKFAIKACLDNRNPEAFQPKMGKGLCAAKLRHSGKGNCK-----RSGCLKN 167
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 SNV-GVK-CLQNKQPQ-WQKSKPKVCNGRLEGNEICDCGTEAOGGPASCCOFTVCUKD 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 168 YCECYEAKIMCSITCK-----IACKNYEESPERKMLMSTPHMEPGDFESSHLYSPAKFS 223
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 219 GAKCY--KGLCKKDCQILQSGVECRP-KAHPECDIA-----ENCNGSSPE--C 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 224 GPP-----KLKNRQAFCSTISWEV--VENTACALLAGEEAEQEHCSPSLAEQMILEBFG 276
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 GPDTTLINGLSCKNNKFCYDGDCHDLDACESVFGSGRNAPACPYEEIOSQS--DRFG 319
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

	Query Match	5.9%	Score 94.5;	DB 1;	Length 577;
	Best Local Similarity	20.5%;	Pred. No. 2.3;		
	Matches	62;	Conservative	37;	Mismatches 110; Indels 93; Gaps 17;
Qy	14	CIDNCGAREUKALHLLPQQDDQSFPPOSEJPKPMTTLVGRLLPPYPAKLNLTQVDNCALP	73		
		: : :	:	:	:
Db	76	CYSNCSPPKQILML-----FPNLYPKKEIT-----LEAFAVIVTON-----	112		
		: : :	:	:	:
Qy	74	SAVNGAAFPSPGALQGPPKITLSGYCDCFSGGDFCN-----SCSNNLRHEL	120		
		: : :	:	:	:

Db 113 -----LALSIGISYDDPKK-----CQSESTCIMPEVVOVSGVKTFFSSCLSRSTQNF 161
QY 121 ERFAKAIACLDNRNEAFQPMGKGRGLGAALKRHSKCNCCK-----RSGCLKN 167
Db 162 SNV-GVK-CLQNKPO-MQKSPKPCVGNRLENEICDCGTEAQCGPASCDFRTCVLKD 218
QY 168 YCEYEAKIMCSSICKC-----TACKNYESPSPKMLMSTPHMYEPDGFESSHYLSPAKFS 223
Db 219 GAKCY--KGLCCCKDCQILQSGVECRP-KAHPECIDIA-----ENCGSSPE--C 261
QY 224 GPP-----KLKRNROAFSCISMEV--VEATCACLLAQGEAEQEHGSPSLAEQMILEEFG 276
Db 262 GPDITILGLSCKNNKFCYDGDCHDLARCESVFGKGRNAPFACYEIQSQS--DREG 319
QY 277 RC 278
Db 320 NC 321

RESULT 14
US-10-155-881-33654
; Sequence 33654, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 33654
; LENGTH: 298
; TYPE: PR1
; ORGANISM: Zea mays
US-10-155-881-33654

Query Match 5.8%; Score 92.5; DB 6; Length 298;
Best Local Similarity 23.7%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCDFSSGDFCNS-CSCNNLRHELRFKAIKACLDNRNEAFQPMGKGRGLGAALKR 152
Db 53 TCGKQCPCLSNCTCCEKYGC-----PKICKNRF----- 81
QY 153 HSKGNCNKRSGCLKNYCEYEAKIMCS-SICK-C-IAC-----KNYE 191
Db 82 --RGCHCAKSQCRSQCPFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 133.

RESULT 15
US-10-155-881-11197
; Sequence 11197, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 11197
; LENGTH: 580
; TYPE: PR1
; ORGANISM: Zea mays

US-10-155-881-11197

Query Match 5.8%; Score 92.5; DB 6; Length 580;
Best Local Similarity 23.7%; Pred. No. 3.5;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCDFSSGDFCNS-CSCNNLRHELRFKAIKACLDNRNEAFQPMGKGRGLGAALKR 152
Db 335 TCGKQCPCLSNCTCCEKYGC-----PKICKNRF----- 363
QY 153 HSKGNCNKRSGCLKNYCEYEAKIMCS-SICK-C-IAC-----KNYE 191
Db 364 --RGCHCAKSQCRSQCPFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 415

Search completed: July 11, 2002, 08:14:15
Job time: 39 sec

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:45:13 ; Search time 15.3956 seconds
(without alignments)
1842.060 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGGAQMLCIDNCGA.....GRCLSQLIHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	396	24.8	429	2 T23152	hypothetical prote
2	363	22.7	571	2 B84585	hypothetical prote
3	357	22.3	603	2 T08955	hypothetical prote
4	232	14.5	658	2 F71410	hypothetical prote
5	115	7.2	1737	2 T00209	MEGF8 protein - hu
6	100	6.3	775	2 T21436	hypothetical prote
7	97	6.1	497	2 T27827	hypothetical prote
8	96.5	6.0	1790	1 MMFR81	laminin beta-1 cha
9	96	6.0	769	2 S55524	male-specific leth
10	96	6.0	3635	2 T10053	laminin alpha 5 ch
11	95	5.9	1096	1 S61917	protein kinase C (
12	95	5.9	1428	2 T08852	lustrin A - Califo
13	94	5.9	716	2 T26998	hypothetical prote
14	94	5.9	1819	2 A71928	cag island protein
15	93	5.8	1816	1 S68960	laminin alpha-4 ch
16	92	5.8	798	2 A28193	integrin beta-1 ch
17	91.5	5.7	383	2 D88633	protein F56B3.2 [i
18	91	5.7	1927	2 G54585	cag pathogenicity
19	91	5.7	2907	2 A57278	fibrillin-2 precur
20	90.5	5.7	735	2 G02937	fertilin beta - cr
21	89	5.6	574	2 B88465	protein B0244.8 [i
22	89	5.6	1786	1 MMWSB1	laminin beta-1 cha
23	88	5.5	625	2 F84706	probable vacuolar
24	88	5.5	798	2 B28193	integrin beta-1* c
25	88	5.5	1786	1 MMHUB1	laminin beta-1 cha
26	87	5.4	290	2 G72858	AcOrf-70 protein -
27	87	5.4	501	2 S56163	tumor necrosis fac
28	87	5.4	652	2 T38704	glycyl tRNA synthe
29	87	5.4	902	2 T01127	curly leaf protein

. ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match 24.8%; Score 396; DB 2; Length 429;
Best Local Similarity 38.9%; Pred. No. 1.2e-25;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YDCDFSSGDFCNSCSC-----NNLRHELRFKAIRACLDNPEAPQPKMKGRLGAKLR- 152

Db 182 YDCDFANGFCDNCCKDCHNNIEYDSQSKAIROSLERNPNAPKPKIGIARGGTIDIER 241

Qy 153 -HSKGCNCRSGCLKNYCEYAEKIMCSICKIACKN-----YEESE---PERKML 199

Db 242 LHQKGCCHKSGCLKNYCEYAEKVPCTDRCKCKGQCTETRYMTRYNKSGGAVSNTNAL 301

Qy 200 MS-----TPHYMEPGDFESSHY---LSPAKFSGPPKLRKNRQAFSCISW-----EVV 243

Db 302 MSLTNASTATPD-SGPGSVVTDEHGDDYEDMLLSHKPKVEMDPRRF---PWYMTDEWV 357

Qy 244 EATCACLAAQGEA-----EQEHCSPSLAEQMILLEEGRCLSQIL 283

Db 358 EATMCWVAQAEALNYEKVQTEDEKLNMEKLVLEFRGCRLEQMI 403

RESULT 2

B84585

hypothetical protein Atg20110 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84585

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84585
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
A:Cross-references: GB:AE002093; NID:g4580462; PIDN:AAD24386.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g20110
A:Map position: 2

Query Match 22.7%; Score 363; DB 2; Length 571;
Best Local Similarity 32.4%; Pred. No. 9.5e-23;
Matches 81; Conservative 39; Mismatches 76; Indels 54; Gaps 8;
QY 36 SSFQPSLPKMTLVGRLL-----PVPKLNLTQVD-NGALPSAVNGAAF 81
DB 46 SSMASIPFIVTVRPIITTSQAPTVATPIPPPOSQIILHVRIRPRPESPMR 105
QY 82 PSGPALQGPVK-----ITLSGYCDCFSGDFCNSCSC-----NNLRHELEPRKAIK 127
DB 106 PAGETRDGTQKKKQCKNCKHSRCLKLYCECFASGTGDCGNCVNCNVEPARQAVE 165
QY 128 ACIDRNPEARQPKMGKRLGA-----LHSGKCNCKRSGLKKNYCEYEAKI 176
DB 166 STLERNPAPRPKTAASPHGGRDNREVGDMVLMARHNGKCHCKSKLKYCECFQANI 225
QY 177 MCSICKIACKNVEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSGP 225
DB 226 LCSENCKLCKNFESEVRQSLFHGHEHSHNLAYLQHANAAITGAGSSGFAS-----APP 281
QY 226 PKLRKNQAF 235
DB 282 PKRRKGQEIF 291

RESULT 3
T08955
hypotheical protein F19B15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08955
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16519
A:Accession: T08955
A:Molecule type: DNA
A:Residues: 1-603 <BEV>
A:Cross-references: EMBL:AL078470; GSPDB:GN00062; ATSP:F19B15.30
A:Experimental source: cultivar Columbia; BAC clone F19B15
C:Genetics:
A:Gene: ATSP:F19B15.30
A:Map position: 4
A:Introns: 107/2; 144/2; 203/3; 321/3; 356/2; 388/1; 522/1
Query Match 22.3%; Score 357; DB 2; Length 603;
Best Local Similarity 35.2%; Pred. No. 3.2e-22;
Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;
QY 36 SSFQPSLPKMTLVGRLLPVPKLNLTQVDNGALPSAVNGA----APPSPALQGP- 90
DB 60 SSVFSTIRP-GMTAIGQVTVQVRPTLPMTATMSNPSPQSOIVNAPTRHPSPKARGPR 118
QY 91 PKI-----TSLGYCDCFSGDFCNSCSC-----NNLRHELEPRKAIK 127
DB 119 PNVEGRDGTQKKKQCKNCKHSRCLKLYCECFASGTGDCGNCVNCNVEPARQAVE 178
QY 128 ACIDRNPEARQPKMGK-----LGAALKL--RHSKCNCKRSGLKKNYCEYEAKI 176
DB 179 ATLERNPAPRPKTAASPHGGRDNREVGDMVLMARHNGKCHCKSKLKYCECFQANI 238
QY 177 MCSICKIACKNVEESPERKMLMSTPH-----YME-----PGDFESSHYL-SPAKFS 223
DB 239 LCSENCKLCKNFESEVRQSLFHGHEHSHNMAYLQOANAAITGAVGSSGFAPSPA--- 295

QY 224 GPPKLRKNRQ 233
DB 296 --PKRRKGQE 303

RESULT 4

F71410
hypotheical protein : Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C:Accession: F71410
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D. P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G. avanach, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B. Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; C.; Chalwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: F71410
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-658 <BEV>
A:Cross-references: GB:Z97337; NID:g2244829; PID:e326824; PID:g2244834
C:Genetics:
A:Map position: 4COP9-4G3845

Query Match 14.5%; Score 232; DB 2; Length 658;
Best Local Similarity 39.5%; Pred. No. 1.1e-11;
Matches 51; Conservative 11; Mismatches 43; Indels 24; Gaps 5;

QY 82 PSGPALQ-----GPPKITLSGYCDCFSGDFC-NSCSC-----NNLRHELEPRKAIKACLD 131
DB 354 PVEPALQELNLSPKK---KSYCEFAAGVYCEPCSCIDCFNKPITHEDVVLATRKQIES 410
QY 132 RNPEAFQPKMGKRLGA-----AKLRHSGKCNCKRSGLKKNYCEYEAKIMCSS 180
DB 411 RNPLAPAPKVRNSDSVQETGDDASKTPASARHKRCNCKKSNCKLKYCECYGGVGCISI 470
QY 181 IKCIACKN 189
DB 471 NCRCEGCKN 479

RESULT 5

T00209
MEGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O. Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like moti
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1737 <NAK>
A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308
A:Experimental source: brain; clone HG1392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Query Match 7.2%; Score 115; DB 2; Length 1737;
Best Local Similarity 23.1%; Pred. No. 0.2;
Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGAREKALHLLPQYDDQSSFPQSELPKPRMTTLVGRLLPVPKLNLTQVDNGALPSAV 76
DB 177 NCGAREKALHLLPQYDDQSSFPQSELPKPRMTTLVGRLLPVPKLNLTQVDNGALPSAV 76

```

Db 711 NCNA-----WLPDPTLRAS-----VGPPMEESVAHAAGVSRILYISGGFGGVALGRL 759
QY 77 NGAAPFSGPA--LQGPPIKTLGYCD-CFS---SGD-----FCNSCSNNL---RHELER 122
Db 760 LALTLPDPCLRLSSPEACNQSGACTWCHGACLSGDAQHRLGCGGSPCPMPSPECCR 819
QY 123 FKAIAKACLDNRNPEAFOPKMGKGRGLGAALKRHSKGC-NCKRSOCL-----KNYC----- 169
Db 820 LRTCSECLARHPTLQPGDGE-----ASTPRCKWCTNCPGACIGRNGSCTSENDCRINQ 874
QY 170 -----ECYEAKI-----MCSSICKCIACKNYEESPE-RKWLMTSTPHY----- 205
Db 875 REVFWAGNCEAAACGAADCEQCTGCKMWTQFKRTGTETRLISVQPTDYDMTCFSHSL 934
QY 206 -MEPGDFESSHYLSPAKFGSPKLRNQAFAFSCISWVVEATCACLLAOGAEAEQHCSP 264
Db 935 NVSPMPVSS---PPLPCTTPCHLLPN-----CTS-----CLDSKGADGGWQHCVW 977
QY 265 SLAEOQMLEEFGRCLS 280
Db 978 SSSLQ-----QCLS 986

RESULT 6
T21436
hypothetical protein R06A4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21436; T23953
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21436
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-775 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB04199.1; GSPDB:GN000020; CESP:R06A4.7
A:Experimental source: clone F26H11
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19822
A:Accession: T23953
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-775 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CAB05589.1; GSPDB:GN000020; CESP:R06A4.7
A:Experimental source: clone R06A4
C:Genetics:
A:Gene: CESP:R06A4.7
A:Map position: 2
A:Introns: 223/1; 414/3; 464/2; 537/1; 601/2; 655/1; 679/1; 749/2

Query Match 6.3%; Score 100; DB 2; Length 775;
Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 32; Conservative 14; Mismatches 41; Indels 60; Gaps 7;

QY 89 GPPKITLGYCDGCFSSGDFCNSCSNNLRHELERFAIKACLDNRNPEAFQPKMGKGRGLGA 148
Db 537 GPCNATAEN-CACRENGVCSYCKCD-----INCSQRF----- 569
QY 149 AKLRHSGCNCKRSGCLKNYCEYEAIKMCS-----SICKC-----IACKNYEESPERK 197
Db 570 -----GCNCAAGQCYTRACQCYRANWCNPMTCNWKCDATDSNLIKRNFGWT---R 619
QY 198 MLMSTPHYMEPGDFESSHYLSLPAKFSG 224
Db 620 MI-----QKRTYCGPSKIAG 634

RESULT 7
T27827
hypothetical protein ZK287.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

```

```

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27827
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27827
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-497 <WIL>
A:Cross-references: EMBL:Z70757; PIDN:CAA94805.1; GSPDB:GN000023; CESP:ZK287.1
A:Experimental source: clone ZK287
C:Genetics:
A:Gene: CESP:ZK287.1
A:Map position: 5
A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2

Query Match 6.1%; Score 97; DB 2; Length 497;
Best Local Similarity 26.6%; Pred. No. 1.8;
Matches 37; Conservative 15; Mismatches 51; Indels 36; Gaps 10;

QY 82 PSGALQGPPIKTLGYCDGCFSSGDFCNS-CSNNLRHELERFAIKACLDNRNPEAFQPK 140
Db 63 PAKPAKNACAK---SERCDLRIKGCCKLACKQI-----C--KNKEA--PK 103
QY 141 MGKGRGLGAALK-RHSKGCNCK--RSGCLKNYCEYEAIKMCSICKICIA-CKNYEESPER 196
Db 104 K-----LAKVAKPTSGCQCAKGGKQCVKKEACRTVYGCASCKCGGDCINGASK--- 154
QY 197 KMLMSTPHYMEPGDFESSH 215
Db 155 ---FSVPKHVQNCFLHKK 170

RESULT 8
MMFFBI
laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; SI4462; B28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988
A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals
A:Reference number: A28783; MUID:88210471; PMID:3365769
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain r
A:Reference number: SI4462
A:Accession: SI4462
A:Molecule type: mRNA
A:Residues: 1-667,'L',669-725,'VT',728-947,950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lambl
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
A:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:
A:Description: interact with cells and with other basement membrane proteins to promo
A:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
F:27-288/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LE01>
F:357-417/Domain: laminin-type EGF-like homology <LE02>
F:420-477/Domain: laminin-type EGF-like homology <LE03>
F:480-528/Domain: laminin-type EGF-like homology <LE04>

```

F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:885-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell adhesion #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>
F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding site
F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
Best Local Similarity 21.0%; Pred. No. 7.2;
Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;
QY 94 TLSGYCDCFSSGDF-----CNSCSC-----NNLRHELERFAKKAACLDNRNPEAFQPKM 141
Db 1021 TTGDHCELCKDGFQDALQNCQCEDFLTGNTNTIAHCDRFTGCPCPL-----PNV 1072
QY 142 KGRLGAARLHRSK-----GCNCRSGCLKNYCEYAKIMCSTICKCIACK-----N 189
Db 1073 QGVRCDOCAENHWHKIASGEGCESCNCDPIGALHEQCNSYTGCCCKPFGGRACNCOAH 1132
QY 190 YEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLRKNRQAFSCISWEVVEATCAC 249
Db 1133 YWGNPNEKQPCCECDQFGAADFQC-----DRETGNVCHEGIGGYKCN 1175
QY 250 LLAOGERAEQHCSP 264
Db 1176 ECARGYIGQFPHCSP 1190

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
Best Local Similarity 21.0%; Pred. No. 7.2;
Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TLSGYCDCFSSGDF-----CNSCSC-----NNLRHELERFAKKAACLDNRNPEAFQPKM 141
Db 1021 TTGDHCELCKDGFQDALQNCQCEDFLTGNTNTIAHCDRFTGCPCPL-----PNV 1072
QY 142 KGRLGAARLHRSK-----GCNCRSGCLKNYCEYAKIMCSTICKCIACK-----N 189
Db 1073 QGVRCDOCAENHWHKIASGEGCESCNCDPIGALHEQCNSYTGCCCKPFGGRACNCOAH 1132
QY 190 YEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLRKNRQAFSCISWEVVEATCAC 249
Db 1133 YWGNPNEKQPCCECDQFGAADFQC-----DRETGNVCHEGIGGYKCN 1175
QY 250 LLAOGERAEQHCSP 264
Db 1176 ECARGYIGQFPHCSP 1190

RESULT 9
S55554
male-specific lethal 2 protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S55554
R:Zhou, S.; Yang, Y.; Scott, M.J.; Pannuti, A.; Fehr, K.C.; Eisen, A.; Koonin, E.V.; Foul
EMBO J. 14, 2884-2895, 1995
A:Title: Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex
A:Reference number: S55554; MUID:95317307; PMID:7796814
A:Accession: S55554
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-769 <ZHO>
A:Cross-references: EMBL:248443; NID:9872110; PIDN:CAA88358.1; PID:9872111
C:Genetics:
A:Gene: FlyBase:msl-2
A:Cross-references: FlyBase:FBgn0005616
A:Introns: 641/1

Query Match 6.0%; Score 96; DB 2; Length 769;
Best Local Similarity 25.74; Pred. No. 3.3;
Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;
QY 121 ERFAIKAC-----LDRNPEAFQPKMKGRLGAARLHRSK-----CNCKRSRG 163
Db 471 KRTLRAKQAAKLEVPSEVTKVQSGK-GA--LRRIRGKDKKEKVPKPKRCRCSIG 527
QY 164 -----CLKNYCEYAKIMCSTICKCIACKN-----YEES-----PER 196
Db 528 SSNTLTTCRNSRCFCYKSYNSCAG-CHVCCKNPKHKEDYVESDEDDLEDFEMPKDYPEP 586

QY 197 KMLMSTPHYMEPGDFESSHYLSPAKFSGPPKL 228
Db 587 MTQSEEPVVAEPROEENS--MAPDSSAPISL 616
RESULT 10
T10053
laminin alpha 5 chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C:Accession: T10053
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
submitted to the EMBL data Library, November 1997
A:Reference number: Z16923
A:Accession: T10053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3635 <MIN>
A:Cross-references: EMBL:U07501; NID:g2599231; PID:g2599232
C:Genetics:
A:Gene: Lama5
C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like h
C:Keywords: basement membrane; cell binding; extracellular matrix
F:1888-1939/Domain: laminin-type EGF-like homology <LEG>
F:1942-1970/Domain: EGF homology <EGF>

Query Match 6.0%; Score 96; DB 2; Length 3635;
Best Local Similarity 18.0%; Pred. No. 17;
Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

QY 14 CIDNCGARELKALHLLPQYDQSQSFPOSELPKMTTLVGRLLPVPKLNILITVDNGLP 73
Db 117 CLERFGPRTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIIVSVLNGR-P 161
QY 74 SAYNGAARF-----SGPALOGP-----PKITLSGY 98
Db 162 GALNFYSYPLLRDFTKATNIRLPLRNTLLGLHLMGRALRDPVTVRRYYYSIKDISIGGR 221
QY 99 CDCFSFGDFC-----NSC-----SC 113
Db 222 CVCHGHADVCDKADPLDPLRLOCAQHNTCGGSCDRCPGPNQPKPATTTSDANECQSC 281
QY 114 NLRH-----ELERFAIKACLDNRNPEAFQPKMKGRLGAARL---RHSKGCNCKRS 162
Db 282 NCHGHAYDYDPEVDR-----RNASQNDNVYQG--GGVCLDCQHHHTTGINGER-- 329
QY 163 GCLKNY-----CEYEAKI---MCSSTI---CK-----CIAC-KNY 190
Db 330 -CLPGFFRAPDQPLDSPHVCBPCDC-ESDFTDGTCEDLTGRCYCRPNFTGELCAACAGY 387
QY 191 EESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLRKNRQAFSCISWEVVEATCAC 250
Db 388 TDRPHCYPLPSFPHN-----DTREQVLPA-----GQIVNCDCAA 422
QY 251 LAOGEAEQEQ-----HC-----SPSLAEQMLIEFGR 277
Db 423 GTGNCACKRDPRLGRVCVKPNFRGAHCELCALPGFHGSPCHPCQCSGPGVANSIDPESGQ 482
QY 278 CL 279
Db 483 CM 484

RESULT 11
S61917
protein kinase C (EC 2.7.1.1-) PKCA - Aspergillus niger
C:Species: Aspergillus niger
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S61917
R:Morawetz, R.; Lendenfeld, T.; Mischak, H.; Muehlbauer, M.; Gruber, F.; Goodnight, J
Mol. Gen. Genet. 250, 17-28, 1996
A:Title: Cloning and characterisation of genes (pkcl and pkca) encoding protein kinas
A:Reference number: S61917; MUID:96158841; PMID:8569684

A:Accession: S61917
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1096 <MOR>
A:Cross-references: EMBL:U10549; NID:9507899; PIDN:AAA97433.1; PID:9507900
C:Genetics:
A:Gene: pkca
A:Introns: 145/1; 234/1; 428/1; 560/1; 698/1; 874/2; 941/3; 1077/1
C:Superfamily: yeast protein kinase C; protein kinase C zinc-binding repeat homology; protein kinase C; duplication; phospholipid binding; phosphotransferase; serine/threonine kinase
F:460-507/Domain: protein kinase C zinc-binding repeat homology <KZ1>
F:528-577/Domain: protein kinase C zinc-binding repeat homology <KZ2>
F:769-1030/Domain: protein kinase C zinc-binding repeat homology <KIN>
F:777-785/Region: protein kinase ATP-binding motif

Query Match 5.9%; Score 95; DB 1; Length 1096;
Best Local Similarity 22.6%; Pred. No. 5.8;
Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;

QY 66 QVDNGALPSAVN-----GAAPFSGPALQPPKITYLSGYDCDFSGDFCNCSNCLRLHEL 120
DB 359 KMEHGAPGRPTAGGAGSAHSGPG--CPGAAPSGQYDGSAGAPAGQVWIDSW-FAL 415
QY 121 ERFKAI-----KACLDNPEAFQPMGKGRIGRAKLR-----HSKG-----C 157
DB 416 EPVGRYILSMSPAKOLKDRPPF---DIGLNQGAVRQKKEVHEKQGHKFVTOQFYNNIM 471
QY 158 NCKRSGCLKNYC---ECYEAKIMCSSIC-----KCIACKNYEESPERKMLMSTHYMEP 208
DB 472 RCALCGDFLUKYAAGQCADCKTCHHKYKPVVTKISKANYETDP----- 517
QY 209 GDFESSHYLSPAKFSGPPKLRNROAFSCISWEVVEATCACLAAQGEAEQHCSPSLAE 268
DB 518 -DEKINHRIHPREG-----FSNIS-----ANWCC-----HCG----- 545
QY 269 QMILBEFR-----CLSQILHIEFKSKGLKIE 295
DB 546 --YLLPFGRKNAKRCCTEGLTCHSHCTHLVDFPCGMSME 582

RESULT 12

T08852

lustrin A - California red abalone
C:Species: *Haliotis rufescens* (California red abalone)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08852
R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E.
J. Biol. Chem. 272, 32472-32481, 1997
A:Title: Molecular cloning and characterization of lustrin A, a matrix protein from shell
A:Reference number: 216496; MUID:98070424; PMID:9405458
A:Accession: T08852
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1428 <SHE>
A:Cross-references: EMBL:AF023459; NID:g2723361; PIDN:AAB95154.1; PID:g2723362
A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial
C:Superfamily: antileukoproteinase repeat homology
C:Keywords: extracellular matrix; extracellular protein
F:1382-1426/Domain: antileukoproteinase repeat homology <ALP>

Query Match 5.9%; Score 95; DB 2; Length 1428;
Best Local Similarity 22.8%; Pred. No. 7.7;
Matches 56; Conservative 21; Mismatches 81; Indels 88; Gaps 14;

QY 74 SAVNGAAPFSGPAL-----QGPKYILSGYDCDFSS-----GDFCN-----S 110
DB 409 SCVGAACPRNTACFAAPSGSAV-----C-CYTSPPRPPEPPSPPTGDPCEPGVNVN 462
QY 111 CSCNNLR-----HELERFKAIKACLDNPEAFQ-----KMGKGRIGRAAKLRHSGKNCCKRS 162
DB 463 CTAGTCRLVEDCRIGCCPAVPCIDRDPLPPPPDVCPVGTVPVGA-----DLKQLYCGRG 518
QY 163 G--CLKN-YCECYEA-----KIMCS-SICKCIA 186

DB 519 GRRCPWNTYCVIHPADRYAVVCCFGSPASAIAPTAPGVDPDCEPGVNVNCTIGVCRLVA 578
QY 187 CKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEAT 246
DB 579 NCDYWPFCARPCTCVD--HSPEP-----SLNCTIGDPALNGKLEEFSCVGGRLCPLN 627
QY 247 CACLLA 252
DB 628 TACLAA 633
RESULT 13
T26998
hypothetical protein Y48B6A.6 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26998
R:Wall, M.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20297
A:Accession: T26998
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-716 <WIL>
A:Cross-references: EMBL:AL110490; NID:el542263; PIDN:CAB54442.1; CESP:Y48B6A.6
C:Genetics:
A:Gene: CESP:Y48B6A.6
A:Introns: 38/3; 196/3; 437/2; 460/3; 518/1; 617/3; 673/3

Query Match 5.9%; Score 94; DB 2; Length 716;
Best Local Similarity 25.0%; Pred. No. 4.6;
Matches 58; Conservative 33; Mismatches 95; Indels 46; Gaps 10;

QY 22 ELKALHLLPQYDDQ-----SSFPQSELKPKMTTLVGRL---LPVPAKLNLITQVD- 68
DB 378 EVKALEAQEASDDRPASPTPIRDSSLPPPPPPKPTPLAIRAGPIPTPQLLEMIHQEDC 437
QY 69 ---NGALPSAVNGAAPPSPALOGPPKITYLSGYDCDFSSDFCNCSNCLRLHELRFKAI 126
DB 438 SIRPSPTSVSHGSRPQSPAPVKPSPVTVS-----PLGLLCDP-----NL--SIEKPEEM 485
QY 127 KACLDNPEAFQPMGKGRIGRAAKLRHSGKNCCKRSGCLKNYCEYEAKIMCSSICKCIA 186
DB 486 KTEDIKPVETAPAPVDEAELNDALDRNK---INATCL-----TKIFFSLVSKYQ 534
QY 187 CKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCI 238
DB 535 NCSFDPLNDKRSNGENVPLKMNK---KISLYAEFEFS-----RKQIQYFSGI 578

RESULT 14

A71928

cag island protein - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD06047.1; PID:g415

A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 5.9%; Score 94; DB 2; Length 1819;
Best Local Similarity 21.4%; Pred. No. 12;
Matches 63; Conservative 39; Mismatches 113; Indels 80; Gaps 12;

QY 22 ELKALHLLPOYDDQSFPOSELPKPK-----MTLVCRLLPVPAK-----LNLIQVDNG 70
DB 568 DILAKESLAKYDCASQAKTEAEKCEKLLTPPEAKKLLLEAEKESVKAYLDCVSKAQTE 627
QY 71 ALPSAVNGAFAFPGPALOGPPKTTLSGYGDCFSFGDFSCNCCNNRHELERFKAKL 130
DB 628 AEKKECEKLLTPPAKKLLEAKSVAYLDCVSKA-----KNAERKECEKLL- 675
QY 131 DRNPEAFQPMKGRIGAAKLHRSKQNC-----KRSGLCKNYCECYEAKIMCSCSIC 182
DB 676 --TPE-----AKLLENQALDCLKNAKTDEERKECKLDPKDLQKVLAKESV 721
QY 183 K-----CIA-CKNYEESPERKMLMSTHYMEPCGDFESSHYLSYPAKFSGPPKLRKNQAFSC 237
DB 722 RYVLDVCSRAKNAEAKKECKLL-TP-----EARKLLEAKKS----- 758
QY 238 ISWEVVEATCACLAQGEAEQHCSPSLAEQ--MILEEFGRLCSOILHIEFKSK 290
DB 759 -----VKAYKDCVLRARNEKEKECEKLLTPPEAKLLEESKSVKAYLDCVSKAK 808

RESULT 15
S68960
laminin alpha-4 chain precursor - human
N:Alternate names: laminin Ah
C:Species: Homo sapiens (man)
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 16-Jun-2000
C:Accession: S68960; S65926; S49149; S40150; I53516
R:Richards, A.; Al-Imara, L.; Pope, F.M.
Eur. J. Biochem. 238, 813-821, 1996
A:Title: The complete cDNA sequence of laminin alpha-4 and its relationship to the other
A:Reference number: S68960; MUID:96300249; PMID:8706685
A:Accession: S68960
A:Molecule type: mRNA
A:Residues: 1-1816 <RIC>
A:Cross-references: EMBL:X91171; NID:g1212962; PIDN:CAA62596.1; PID:g1212963
A:Experimental source: Tissue type heart
R:Ilvanainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.
FEBS Lett. 365, 183-188, 1995
A:Title: Primary structure and expression of a novel human laminin alpha-4 chain.
A:Reference number: I53516; MUID:95300971; PMID:7781776
A:Accession: S65926
A:Molecule type: mRNA
A:Residues: 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816 <IIIV>
A:Cross-references: EMBL:S78569; NID:g1042081; PIDN:AB34635.1; PID:g1042082
R:Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.; Lloyd, J.C.; Pope, F.M.
submitted to the EMBL Data Library, December 1993
A:Description: Localisation of the gene (LAMA4) to chromosome 6q21 and isolation of a pa
A:Reference number: S49149
A:Accession: S49149
A:Molecule type: mRNA
A:Residues: 236-1816 <RI2>
A:Cross-references: EMBL:X76939; NID:g509805; PIDN:CAA54258.1; PID:g509806
R:Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope, F.M.
submitted to the EMBL Data Library, February 1993
A:Description: Isolation of a partial cDNA encoding a protein homologous to laminin A.
A:Reference number: S40150
A:Accession: S40150
A:Molecule type: mRNA
A:Residues: 1403-1541, 'S', 1543-1816 <RI3>
A:Cross-references: EMBL:X70904; NID:g437804; PIDN:CAA50261.1; PID:g437805
C:Genetics:
A:Gene: GDB:LAMA4; LAMA3
A:Cross-references: GDB:203904; OMIM:600133
A:Map position: 6q21-6q21
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like H

C:Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glyco
F:1-24/Domain: signal sequence #status: predicted <SIG>
F:25-1816/Product: laminin alpha-4 chain #status: predicted <MAT>
F:82-129/Domain: laminin-type EGF-like homology <LE1>
F:132-184/Domain: laminin-type EGF-like homology <LE2>
F:187-238/Domain: laminin-type EGF-like homology <LE3>
F:241-265/Domain: laminin-type EGF-like homology #status: atypical <LE4>
F:717-719/Region: cell attachment (R-G-D) motif
F:862-1031/Domain: laminin G repeat homology <LG1>
F:1068-1223/Domain: laminin G repeat homology <LG2>
F:1252-1398/Domain: laminin G repeat homology <LG3>
F:1488-1636/Domain: laminin G repeat homology <LG4>
F:1665-1816/Domain: laminin G repeat homology <LG5>
F:104,215,308,458,524,550,571,574,631,639,735,751,754,780,803,1086,1281,1359,1411/Bin
F:266,269/Disulfide bonds: Interchain #status: predicted

Query Match 5.8%; Score 93; DB 1; Length 1816;
Best Local Similarity 23.8%; Pred. No. 14;
Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

QY 2 VICQLKGAQMLCIDNCG-----ARELKALH---LLPOY---DDQSFPOSELPKPMITLV 51
DB 80 VPCDCNGSNE-CLDGSYCVHCQRNTTGEHCCKLDYIGDSIRGAPQFCQPCP----- 133
QY 52 GRLLPVPKLNLTQV--DNGALPSAVN-GAAPPS-----GPAQQGPPKITLSGYCDCFSS 104
DB 134 ---CPLPHLANFAESCVRKNGAVRCICNENYAGPNCERCAPGYGNPLLI----- 180
QY 105 GDFCNSCSN-----NLRHE--LERFKAIKACLDNRPEAFQ-PKMGKGRIGAAKL-RHSK 155
DB 181 GSTCKKDCSGNSDNLIFEDCDEVGTGQCRNCL-RNTTGFKCKERCAPGYGDIARAKNCA 239
QY 156 GCNCKRSGCLKNYCECYEAKIMCSCSIC-RCI-----ACKNYEESPERKMLMST--P 203
DB 240 VNCGGGPCDSVTGCECLESEGFEPPTGCDKCVMDLTDDLRLAALSIEEGKSGVLSVSGAA 299
QY 204 HYMEPGDFESSHYLSYPAKFSGPPKLRKNQAFSCISWEVVEATCACLAQGEAEQHC 263
DB 300 AHRHVNEINATIIYLLKTLSE-----RENOYALRKIQINNAENTMKLSLLSDVEELVEKENQ 355
QY 264 PSLAEQMILEE 274
DB 356 ASRKGLVQKE 366

Search completed: April 21, 2003, 11:50:39
Job time : 21.3956 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:43:43 ; Search time 27.8114 Seconds
(without alignments)
2185.571 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGAGMLCIDNCGA.....GRCLSQLTHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertibrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	100.0	475	11	O8VIE0 mus musculu
2	1589	99.4	475	11	O8VIE1 mus musculu
3	504	31.5	277	11	O9D571 mus musculu
4	435	27.2	950	5	O9V6Q8 drosophila
5	396	24.8	429	5	O62295 caenorhabdi
6	396	24.8	435	5	O95QD7 caenorhabdi
7	363	22.7	571	10	O9SL70 arabadopsis
8	357	22.3	603	10	O9SZD1 arabadopsis
9	253.5	15.9	356	10	O9LW71 arabadopsis
10	251.5	15.7	601	10	O9CAV1 arabadopsis
11	232.5	14.5	896	10	O9ZS22 glycine max
12	232	14.5	658	10	O23333 arabadopsis
13	223.5	14.0	609	10	O9M679 arabadopsis
14	222	13.9	526	10	O94A12 arabadopsis
15	222	13.9	695	10	O9LE32 arabadopsis
16	222	13.9	695	10	O9LUI3 arabadopsis

17	186	11.6	243	5	O9VMQ3
18	180.5	11.3	593	10	O9LUI5
19	159	9.9	553	10	O94DS2
20	115	7.2	1737	4	O75097
21	101	6.3	769	11	O8VHL6
22	100	6.3	773	5	O17514
23	100	6.3	773	5	O62335
24	99.5	6.2	1704	5	O94446
25	99	6.2	2931	5	O9W2C6
26	97.5	6.1	772	13	O9PU04
27	97	6.1	497	5	O23460
28	97	6.1	984	5	O9Y1P7
29	97	6.1	1106	5	O9GU55
30	97	6.1	1987	5	O9V841
31	96.5	6.0	1026	5	O8SWY0
32	96.5	6.0	4547	5	O9W343
33	96	6.0	773	5	O9VQR1
34	95.5	6.0	448	5	O9NKE0
35	95.5	6.0	1035	5	O9VJU8
36	95	5.9	1428	5	O44341
37	94.5	5.9	631	15	O9IEZ0
38	94	5.9	716	5	O9U2A6
39	94	5.9	1819	16	O9ZLV0
40	93.5	5.8	420	5	P91776
41	93.5	5.8	1981	5	O9VVK7
42	93	5.8	463	11	O8ROC7
43	93	5.8	761	5	O76733
44	93	5.8	4123	4	O75851
45	92.5	5.8	489	11	O9D211

ALIGNMENTS

RESULT 1

O8VIE0	ID	O8VIE0	PRELIMINARY;	PRT;	475 AA.
AC	ORVIE0:				
DT	01-MAR-2002 (TREMBlrel. 20, Created)				
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)				
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)				
DE	Tesmin (Hypothetical 50.6 kDa protein).				
GN	TESM.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BALE/C;				
RA	Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;				
RT	"Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of				
RT	localization throughout spermatogenesis."				
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Strausberg R.;				
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB057423; BAB64935.1; -				
DR	EMBL; BC024377; AAH24377.1; -				
DR	InterPro: IPR005172; CXC.				
DR	Pfam: PF03638; CXC; 1.				
KW	Hypothetical protein.				
SO	SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF79F CRC64;				

Query Match 100.0%; Score 1599; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	MVICQLKGGAGMLCIDNCGARELKALHLLPQYDQSSFPQSELPKPMVTTLVGRLLPVPK 60
Db	181	MVICQLKGGAGMLCIDNCGARELKALHLLPQYDQSSFPQSELPKPMVTTLVGRLLPVPK 240

```

QY 61 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 120
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 300
QY 121 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 180
DB 301 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 360
QY 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 240
DB 361 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 420
QY 241 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 295
DB 421 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 475

RESULT 2
Q8VIE1 PRELIMINARY; PRT; 475 AA.
AC Q8VIE1
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE tesmin.
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057422; BAB64934.1;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Query Match 99.4%; Score 1589; DB 11; Length 475;
Best Local Similarity 99.3%; Pred. No. 7.1e-160;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVICQLKGGQAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
DB 181 MVICQLKGGQAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK 240
QY 61 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 120
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 300
QY 121 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 180
DB 301 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 360
QY 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 240
DB 361 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 420
QY 241 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 295
DB 421 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 475

RESULT 3
Q9D571 PRELIMINARY; PRT; 277 AA.
ID Q9D571
AC Q9D571

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4930509C02, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1;
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;

Query Match 31.5%; Score 504; DB 11; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.2e-45;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVICQLKGGQAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
DB 181 MVICQLKGGQAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPK 240
QY 61 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 120
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGPPTLTSYCDGCFSSGDFCNSCNSNLRHEL 300
QY 121 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 180
DB 301 ERFAIKACLDNRPEAFQPMKGRGLCAALRHSGKCNCRSGCLKNYCEYEAKIMCSS 360
QY 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 240
DB 361 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 420
QY 241 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 295
DB 421 EVVEATCACLAAQEAEQEHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 475

RESULT 4
Q9V608 PRELIMINARY; PRT; 950 AA.
ID Q9V608
AC Q9V608
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

```


RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boltshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Dahle C., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fozler C., Gabriellian A.C., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin G., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobaraj C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003818; AAF58365.1;
 DR FlyBase: FBgn0033846; CG6061.
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 950 AA; 100021 MW; 627C3EA6B44A0A30 CRC64;

Query Match 27.2%; Score 435; DB 5; Length 950;
 Best Local Similarity 38.48; Pred. No. 4e-37;
 Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy 31 QYDQSSPQSELKPMTLYGRLPVPAKLNLTQVDNGALPVAAGAAPSPGALQGP 90
 Db 681 OLPTQSTPIKVEKPLTPPGVKANVPK--PLFEVLKPPATAAAGAVDPLGWTSSR 738
 Qy 91 PKITLSG-----YDCFSFGDFCNSCSC-----NNLRHELRFKAIKACLDNRNPEAFOPK 140
 Db 739 KHCNCKSQCLKLYDCDFANGFCQDCCTCKDCFNLDYEVERAIRSCLDRNPSAFKPK 798
 Qy 141 MCKGRLGAALRHSGCKRGCKNYCEYEAQKIMCSSICKCTACKNYEESPCKMLM 200
 Db 799 ITAPNSGDMRL-HNKGCKNSGCKNYCEYEAQKIMCSSICKCTACKNYEESPCKMLM 852
 Qy 201 STPHMEPGDFESHYSHPKFPKLRNRQ-----AFSCISWEVVEATCACLAAQGE 255
 Db 853 -----VMDSLDLGLMGVEQKKDK-AKNQLNENRANIYFTDDVIEATIMCISRV 903
 Qy 256 EAEQHCSPSLAEQMIIEFGRCLSQIL 283
 Db 904 MHEKONVAMEDMEVMEEMGESLTQII 931

RESULT 5
 ID O62295 PRELIMINARY; PRT; 429 AA.
 AC O62295;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE JCB.6b protein.
 GN JCB.6b.
 OS Caenorhabditis elegans.

Query Match 24.8%; Score 396; DB 5; Length 435;
 Best Local Similarity 38.9%; Pred. No. 2.1e-33;
 Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YDCFSFGDFCNSCSC-----NNLRHELRFKAIKACLDNRNPEAFOPKMGRLGAALR- 152
 Db 182 YDCDFANGFCQDCNCKDCHNNIEYDSQSKAIRQSLEARNPNAPKPKIGIARGGITDIER 241
 Qy 153 -HSKGCNCKRGCKNYCEYEAQKIMCSSICKCTACKN-----YEEES-----PERKML 199
 Db 242 LHQGCCHCKGCKNYCEYEAQKIMCSSICKCTACKN-----YEEES-----PERKML 301
 Qy 200 MS-----TPHYMEPGDFESHYSHPKFPKLRNRQAFSCISW-----EW 243
 Db 302 MSLTNASSTATPD-SGPGSVVTDEHGDDYEDMLLSHKPKVEMDPRPF---PWYMTDEV 357
 Qy 244 EATCACLAAQGEA-----EQEHCSPSLAEQMIIEFGRCLSQIL 283
 Db 358 EAATMCVMAQAEALNYEKVQTEDEKLNMEKLVLRFGRCLEQMI 403

RESULT 6
 ID Q95QD7 PRELIMINARY; PRT; 435 AA.
 AC Q95QD7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE JCB.6a protein.
 GN JCB.6a.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN RA Lighting J.;
 RP Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RL SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z82274; CAB05228.1; -;
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 435 AA; 48365 MW; 10601B22681C06D2 CRC64;

Query Match 24.8%; Score 396; DB 5; Length 429;
 Best Local Similarity 38.9%; Pred. No. 2e-33;
 Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YDCFSFGDFCNSCSC-----NNLRHELRFKAIKACLDNRNPEAFOPKMGRLGAALR- 152
 Db 182 YDCDFANGFCQDCNCKDCHNNIEYDSQSKAIRQSLEARNPNAPKPKIGIARGGITDIER 241
 Qy 153 -HSKGCNCKRGCKNYCEYEAQKIMCSSICKCTACKN-----YEEES-----PERKML 199
 Db 242 LHQGCCHCKGCKNYCEYEAQKIMCSSICKCTACKN-----YEEES-----PERKML 301
 Qy 200 MS-----TPHYMEPGDFESHYSHPKFPKLRNRQAFSCISW-----EW 243
 Db 302 MSLTNASSTATPD-SGPGSVVTDEHGDDYEDMLLSHKPKVEMDPRPF---PWYMTDEV 357
 Qy 244 EATCACLAAQGEA-----EQEHCSPSLAEQMIIEFGRCLSQIL 283
 Db 358 EAATMCVMAQAEALNYEKVQTEDEKLNMEKLVLRFGRCLEQMI 403

RESULT 6
 ID Q95QD7 PRELIMINARY; PRT; 435 AA.
 AC Q95QD7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE JCB.6a protein.
 GN JCB.6a.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN RA Lighting J.;
 RP Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RL SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z82274; CAB05228.1; -;
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 435 AA; 49109 MW; 7629A5C749D25A5D CRC64;

Query Match 24.8%; Score 396; DB 5; Length 435;
 Best Local Similarity 38.9%; Pred. No. 2.1e-33;

RA Pihakaski-Maunsbach K., Marcker K.A., Jensen E.O.;
RT "Cppl", a novel type DNA-binding protein involved in the expression of
TL soybean leghemoglobin c3 gene."
RL Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
DR EMBL: AJ010165; CA009028.1; -
DR InterPro: IPR005172; CXG.
DR Pfam: PF03638; CXG; 2.
SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

Query Match 14.5%; Score 232.5; DB 10; Length 896;
Best Local Similarity 38.3%; Pred. No. 1.1e-15;
Matches 44; Conservative 16; Mismatches 38; Indels 17; Gaps 3;
QY 98 YDCDFSGGDFC-NSCSCNNLRHLEFRKAI-----KACLDNRNPEAFQPKM----- 141
DB 487 YDCDFAGTYCTPCACOGCLNRPVEVETVETKQIESRNPFAFKIVQPTTIDISSHM 546
QY 142 -GGRGLGAALRKSGCKNCKRSGLKNCYCEYAKIMCSCICKIACKNVEESPE 195
DB 547 DDENLTPSSARHKGCKNCKRSCLKCYCEYQANVGCSSGRCGCKNVHGKKE 601

RESULT 12
O23333 PRELIMINARY; PRT; 658 AA.
AC O23333;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 72.1 kDa protein.
GN AT4G14770.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SQ SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staaven M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Welten E., Wambutt R., Weitzenger T., Pohl T.M., Terryn N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kottler P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA Dalseny M., Voet M., Volckaert G., Mewes H.W., Klosternan S.,
RA Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana";
RL Nature 391:485-488(1998).
RN [2]
SQ SEQUENCE FROM N.A.
RA EG Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL: 297337; CAB10256.1; -
DR ENBL: AL161539; CAB78519.1; -
DR InterPro: IPR005172; CXG.
DR Pfam: PF03638; CXG; 2.
KW Hypothetical protein.
SQ SEQUENCE 658 AA; 72129 MW; 5BB99148E5E13A77 CRC64;

Query Match 14.5%; Score 232; DB 10; Length 658;
Best Local Similarity 39.5%; Pred. No. 8.6e-16;
Matches 51; Conservative 11; Mismatches 43; Indels 24; Gaps 5;
QY 82 PSPALQ-----GPKITLGGYDCDFSGGDFC-NSCSC-----NNLRELEFRKAIACLD 131
DB 354 PVEPALQELNLSFPKK---KSYCECFANGYCIPEPCIDCFNKPHEDEVVLAIRKQIES 410

QY 132 RNPEAFQPMRGMGRGLG-----AKLRHSGCKNCKRSGLKNCYCEYAKIMCSC 180
DB 411 RNPEAFAPKVRNSDSVQETGDDASKTPASARHKGCKNCKSKLKYCEYGGVGCISI 470
QY 181 IKCKIACKN 189
DB 471 NCRCEGCKN 479

RESULT 13
Q9M679 PRELIMINARY; PRT; 609 AA.
AC Q9M679;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXG domain containing TSOL-like protein 1.
GN SOL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SQ SEQUENCE FROM N.A.
RP STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TSOL is a novel protein that modulates cytokinesis and cell expansion
RL in Arabidopsis";
RL Development 127:2219-2226(2000).
DR EMBL: AF205142; AAF69125.1; -
DR InterPro: IPR005172; CXG.
DR Pfam: PF03638; CXG; 2.
SQ SEQUENCE 609 AA; 66738 MW; 7E29C055A2423089 CRC64;

Query Match 14.0%; Score 223.5; DB 10; Length 609;
Best Local Similarity 31.7%; Pred. No. 6.2e-15;
Matches 52; Conservative 21; Mismatches 36; Indels 55; Gaps 7;
QY 98 YDCDFSGGDFC-NSCSCNNLRH-----LERFKAICLDNRNPEAFQPKMCKG----- 144
DB 341 YCECFAGTYCTPCACOGCLNRPVEVETVETKQIESRNPFAFKIVQPTTIDISSHM 397
QY 145 RLQ-----AAKLRHSGCKNCKRSGLKNCYCEYAKIMCSCICKIACKN----- 189
DB 398 EVGEDASKTPASARHKGCKNCKSKLKYCEYGGVGCISINCRCEGCKNAFGRKDGSL 457
QY 190 YESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGGPKLRKNQ 233
DB 458 FEQDEENET-----SGTPGTTKTKQ 477

RESULT 14
Q94AI2 PRELIMINARY; PRT; 526 AA.
AC Q94AI2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative DNA binding protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SQ SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,

```

RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full Length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY046019; AAK76693.1; -
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 526;
Best Local Similarity 31.8%; Pred. No. 7.5e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YCDGSSGDFC-NSCSC----NNLRHELRFKAKACLDNRNPEAFQPKM----- 141
DB 245 YCECFAGVYCIIPCSCIDCFNKPHEETVLTATRKQIESRNPFAFAPKVI RNADSIMEAS 304
QY 142 GKRLGAALKRSHKSGCKRSGCLKNYCEYEAKEIMCISICKCIACKN----- 189
DB 305 DDASKTPASARHKGCKNCKKCMKKYCEYQGVGCSMNCRCGCTNVFGRKDGSLVI 364
QY 190 ----YESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
DB 365 MESKLEENQETYEKRIAKIOHNVESKEVQNPSSDQPSPLPPYRHLVVHQPFLSKNR 423

RESULT 15
Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TS01 (Putative DNA binding protein).
GN TS01.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TS01 is a novel protein that modulates cytokinesis and cell expansion
RL in Arabidopsis.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
RL Arabidopsis gene with cysteine-rich repeats.";
DR EMBL; AF204059; AAF69124.1; -
DR EMBL; AF206324; AAF27433.1; -
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 1.1e-14;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YCDGSSGDFC-NSCSC----NNLRHELRFKAKACLDNRNPEAFQPKM----- 141
DB 414 YCECFAGVYCIIPCSCIDCFNKPHEETVLTATRKQIESRNPFAFAPKVI RNADSIMEAS 473
QY 142 GKRLGAALKRSHKSGCKRSGCLKNYCEYEAKEIMCISICKCIACKN----- 189
```

```

DB 474 DDASKTPASARHKGCKNCKKCMKKYCEYQGVGCSMNCRCGCTNVFGRKDGSLVI 533
QY 190 ----YESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
DB 534 MESKLEENQETYEKRIAKIOHNVESKEVQNPSSDQPSPLPPYRHLVVHQPFLSKNR 592

Search completed: April 21, 2003, 11:49:57
Job time : 29.8114 secs
```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:30:23 ; Search time 33.2222 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 NVICLKGGTQMLCIDNSRT.....GRCLSQLHTEFKSKGLKWE 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	100.0	299	21	Human testis speci
2	1251	76.0	295	21	Mouse testis speci
3	470	28.5	950	22	Drosophila melanog
4	446	27.1	147	22	Human testicular a
5	446	27.1	147	22	Human reproductive
6	441	26.8	438	20	Caenorhabditis ele
7	378	23.0	280	22	Novel human diagno
8	278	16.9	53	22	Human colon cancer
9	216	13.1	251	22	Human polypeptide,
10	200	12.1	243	22	Drosophila melanog

11	113	6.9	1981	22	ABR61657	Drosophila melanog
12	109.5	6.6	699	20	AAW81840	Human LFA-1 beta c
13	109.5	6.6	769	9	APB0836	Beta subunit of hu
14	109.5	6.6	769	11	AA07113	Recombinant beta-s
15	109.5	6.6	769	16	AA080108	LFA-1 beta subunit
16	109.5	6.6	769	21	AA03974	LFA-1 CD18 polyep
17	109.5	6.6	793	22	ABG16399	Novel human diagno
18	108.5	6.6	769	13	AA024256	Beta-subunit CD18
19	108	6.6	676	18	AAW35853	Human CD18 for use
20	108	6.6	3011	13	AA021519	Compiled HCV seque
21	106.5	6.5	3635	23	ABR1589	Mouse laminin alph
22	106.5	6.5	3635	23	AAW50357	Mouse laminin-15 a
23	105.5	6.4	695	22	ABR30979	Peptide #3630 enco
24	105.5	6.4	695	22	ABR36170	Peptide #3676 enco
25	105.5	6.4	695	22	AAW58953	Human brain expres
26	105.5	6.4	695	22	AAW69338	Human bone marrow
27	105.5	6.4	695	22	AAW17170	Peptide #3604 enco
28	105.5	6.4	695	22	AAW29661	Peptide #3698 enco
29	105.5	6.4	695	22	AAW04863	Peptide #3545 enco
30	105.5	6.4	695	23	ABG38956	Human peptide enco
31	104	6.3	3011	14	AA031621	Hepatitis C virus
32	103.5	6.3	1725	21	AAW19800	Mouse laminin 2 ma
33	103.5	6.3	1725	21	AAW48451	Mouse laminin 8 po
34	103.5	6.3	1725	23	ABR1593	Mouse laminin 10 s
35	103.5	6.3	1764	10	AAW91672	Primary amino acid
36	103.5	6.3	1776	19	AAW50894	Mouse laminin BI c
37	103.5	6.3	1786	21	AAW19799	Mouse laminin 2 be
38	103.5	6.3	1786	21	AAW48450	Mouse laminin 8 po
39	103.5	6.3	1786	23	ABR1592	Mouse laminin 10 s
40	102	6.2	1788	22	ABR62995	Drosophila melanog
41	101.5	6.2	686	23	AAE18689	HCV-1 NS3/4a mutan
42	101.5	6.2	686	23	AAU76377	Hepatitis C virus
43	101.5	6.2	728	23	AAE18688	NS3/4a mutant conf
44	101.5	6.2	762	22	ABG08032	Novel human diagno
45	101	6.1	473	23	AAW50076	HCV NS3 helicase p

ALIGNMENTS

RESULT 1	
AAW68464	
ID	AAW68464 standard; Protein; 299 AA.
AC	AAW68464;
XX	
DT	25-APR-2000 (first entry)
XX	
DE	Human testis specific factor tesmin SEQ ID NO:5.
XX	
KW	Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW	differentiation regulatory factor; male germ cell regulatory actor;
KW	germ cell differentiation; sterility.
XX	
OS	Homo sapiens.
XX	
PN	WO200004147-A1.
XX	
PD	27-JAN-2000.
XX	
PF	16-JUL-1999; 99WO-JP03859.
XX	
PR	17-JUL-1998; 98JP-0219856.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	(AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PI	Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX	
DR	WPI; 2000-147785/13.
XX	N-PSDB; AA288156.
PT	New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility -
XX
PS Claim 1; Page 50-52; 63pp; Japanese.
XX
CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 299 AA;

Query Match 100.0%; Score 1647; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.8e-141;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDONNYLQSDVPKPMALVGRFLPASTK 60
Db 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDONNYLQSDVPKPMALVGRFLPASTK 60

Qy 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120
Db 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120

Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 180
Db 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 180

Qy 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
Db 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240

Qy 241 CISWEVVEATCACLAAQGEAEKHCSCLAEQMILEEFGRCLSQLHTEFKSKGLKME 299
Db 241 CISWEVVEATCACLAAQGEAEKHCSCLAEQMILEEFGRCLSQLHTEFKSKGLKME 299

RESULT 2
AA68463
ID AA68463 standard; Protein: 295 AA.
XX
AC AA68463;
XX
DT 25-APR-2000 (first entry)
XX
DE Mouse testis specific factor tesmin SEQ ID NO:4.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Mus musculus.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-JP03859.
XX
PR 17-JUL-1998; 98JP-0219856.
XX
PA (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX
DR WPI; 2000-147785/13.
DR N-PSDB; AAZ88155, AAZ88157.
XX
PT New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -
XX
PS Claim 1; Page 47-49; 63pp; Japanese.

CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 295 AA;

Query Match 76.0%; Score 1251; DB 21; Length 295;
Best Local Similarity 76.3%; Pred. No. 2.5e-105;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDONNYLQSDVPKPMALVGRFLPASTK 60
Db 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDONNYLQSDVPKPMALVGRFLPASTK 60

Qy 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120
Db 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 116

Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 180
Db 117 RHELEERFAIKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 176

Qy 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
Db 177 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 236

Qy 241 CISWEVVEATCACLAAQGEAEKHCSCLAEQMILEEFGRCLSQLHTEFKSKGLKME 299
Db 237 CISWEVVEATCACLAAQGEAEKHCSCLAEQMILEEFGRCLSQLHTEFKSKGLKIE 295

RESULT 3
ABB62035
ID ABB62035 standard; Protein: 950 AA.
XX
AC ABB62035;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 12897.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06138.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483232/52.

Nucleic acids encoding 973 human testicular antigen polypeptides,
useful for preventing, diagnosing and/or treating testicular cancer

Claim 11; SEQ ID NO 1409; 766pp; English.

The present invention provides the protein and coding sequences of 973
human testicular antigens, and fragments of their genomic sequences. The
sequences can be used in the treatment of cardiovascular, urinary system,
reproductive system, immune, respiratory, neurological and
gastrointestinal disorders, infections, and particularly cancer,
especially testicular cancers. The present sequence is a protein of the
invention.

Sequence 147 AA;

Query Match 27.1%; Score 446; DB 22; Length 147;

Best Local Similarity 98.9%; Pred. No. 1.3e-32;

Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVICQLKGGTQMLCINSTRRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
|||||
Db 60 MVICQLKGGTQMLCINSTRRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 119
|||||
QY 61 LNLITQOLEGALPSVVGSAFPGSGTLP 88
|||||
Db 120 LNLITQOLEGALPSVVGSAFPGSGTLP 147
|||||
RESULT 5
AAM95330
ID AAM95330 standard; Protein; 147 AA.
XX
AC AAM95330;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 3988.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

ID AAG76158 standard; Protein; 53 AA.
XX
AC AAG76158;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6922.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR N-PSDB; AAH35563.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 8365; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 53 AA;
Query Match 16.9%; Score 278; DB 22; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 53; Conservative: 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 VEATCACLAAQGEAEKHCSCKLABQMILEEFGRCLSQILHTEFKSKGLKME 299
Db 1 VEATCACLAAQGEAEKHCSCKLABQMILEEFGRCLSQILHTEFKSKGLKME 53
RESULT 9
ID AAM93348
AM93348 standard; Protein; 251 AA.
XX
AC AAM93348;
XX
DT 06-NOV-2001 (first entry)
XX
XX

DE Human polypeptide, SEQ ID NO: 2895.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB; AAR94268.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;
Query Match 13.1%; Score 216; DB 22; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVTICQLKGGTQMLCIDNSRTELKALHLVLPQYQDQNNYLQS 41
Db 210 MVTICQLKGGTQMLCIDNSRTELKALHLVLPQYQDQNNYLQS 250
RESULT 10
ID ABB68888
ABB68888 standard; Protein; 243 AA.
XX
AC ABB68888;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.

AA07113 standard; protein; 769 AA.
AA07113;
24-JAN-1991 (first entry).
Recombinant beta-subunit of LFA-1.
Rhinovirus; LFA-1; Mac-1; p150,95; ICAM-1; autoimmune diseases;
graft rejection.
Key Location/Qualifiers
FH Peptide 197..205
AC /label=tryptic peptide L-56A
XX Peptide 258..269
FT /label=tryptic peptide H58
FT Peptide 311..324
FT /label=tryptic peptide P-61
FT Peptide 340..351
FT /label=tryptic peptide H-52
FT Peptide 367..375
FT /label=tryptic peptide L-65
FT Peptide 472..479
FT /label=tryptic peptide P-20
FT Peptide 489..498
FT /label=tryptic peptide P-18
FT Peptide 528..539
FT /label=tryptic peptide L-56b
XX W09010652-A.
XX
XX 20-SEP-1990.
XX 09-MAR-1990; 90WO-US01255.
XX 09-MAR-1990; 90WO-US01255.
XX (DANA-) DANA FARBER CANCER.
XX Springer T, Kishimoto T, Roberts T;
XX WPI; 1990-304991/40.
XX N-PSDB; AAQ06050.
XX Use of recombinant B-sub-unit of human LFA-1 - for treatment of
XX rhino-viral infections
XX Disclosure; fig 1; 2lpp; English.
XX This beta-subunit of the human lymphocyte function-associated
XX glycoprotein (LFA-1) is used, in a compsn. along with the alpha
XX chain of a member of the CD-18 family, for treating or preventing
XX rhinoviral infections. The beta-subunit interacts with ICAM-1
XX and so inhibits cell-viral attachment. The beta-subunits of other
XX human glycoproteins (esp. Mac-1 or p150,95) involved in cellular
XX adhesion can be used also. Antibodies raised against these can be
XX used as assay reagents to detect LFA-1, etc. in tissues, or for
XX purificn. See also AAQ06051-53.
XX
XX Sequence 769 AA;
Query Match 6.6%; Score 109.5; DB 11; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.37;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGTOMLCDINSTRRELKALHLVPOYQDNVYLOSVPKPMYALVGRFLPASTK 60
DB 436 IVTVQVLFPQCECRCDQSDRLS-CH-----GKGFLEGGICRCDDTGYIGKNCCEQTQ 486
QY 61 LNLITQOLEGA-----LPSVYNGSAFPSTGLPGPKKITLAGYCDCEFA----- 103
DB 487 -GRSSQELGSCRDNNNSITCGSLGDCVCGQCLHTSDVPG--KLIIYQYCECDTINCER 543

QY 104 -SGDFCNN-----CNCNNCCNLLHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
DB 544 YNGQVCGGPGRGLCFCKC--RCHPGFEGSACQCERTTEGL--NPRRVECS-GRGR--- 595
QY 153 VKPOHNGCNCRRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCGCPGCKGYIISCAECLKEKGP 638
RESULT 15
AAR80108
ID AAR80108 standard; Protein; 769 AA.
XX
AC AAR80108;
DT 04-MAR-1996 (first entry)
XX
DE LFA-1 beta subunit CD18.
XX
KW CD18; leucocyte function-associated antigen-1; LFA-1; integrin.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FH Peptide 1..22
FT /label= Sig_peptide
FT Region 449..496
FT /label= Repeat
FT /note= "cysteine-rich repeat"
FT Region 497..540
FT /label= Repeat
FT /note= "cysteine-rich repeat"
FT Region 541..581
FT /label= Repeat
FT /note= "cysteine-rich repeat"
FT Region 582..617
FT /label= repeat
FT /note= "cysteine-rich repeat"
FT Domain 701..723
FT /label= Trans
FT /note= "transmembrane domain"
FT Domain 724..769
FT /label= Cyto
FT /note= "cytoplasmic domain"
XX W09528170-A1.
XX
XX 26-OCT-1995.
XX 19-APR-1995; 95WO-US04886.
XX 19-APR-1994; 94US-0229513.
XX (UNIV) UNIV KANSAS.
XX Benedict S, Chan MA, Siahaan TJ, Tibbetts SA;
XX WPI; 1995-373631/48.
XX
XX Changing peptide reactivity via conjugation with a second peptide
XX causes change in conformation of first peptide, pref. LFA-1 and
XX ICAM-1 functional domain derived peptide(s)
XX
XX Claim 11; Page 30-32; 64pp; English.
XX
XX Functional domains derived from LFA-1 alpha subunit CD11a (AAR80107),
XX LFA-1 beta subunit CD18 (AAR80107) and ICAM-1 CD43 (AAR80110) are used
XX as the basis of short, synthetic peptides (AAR80109, AAR80111-24) that
XX modulate ICAM/LFA binding interaction. CD18 is obtd. from human
XX tonsil.
XX
XX Sequence 769 AA;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:48:39 ; Search time 13.5909 Seconds
(without alignments)
647.304 Millioncell updates

Title: US-09-743-237-5

Perfect score:

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQILHTEFKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Loss processing: Minimum Match 0% Maximum Match 100%

Maximum Match 1008
Listing first 45 summaries

Database :

Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*

/cgn2_6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*

/cgn2_6/ptodata/1/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*

5: /cgn2_6/ptodata/1/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query #			DB	ID	Description
	Score	Match	Length			
1	109.5	6.6	769	2	US-08-789-078-1	Sequence 1, Appl
2	109.5	6.6	769	2	US-08-753-633-1	Sequence 1, Appl
3	109.5	6.6	769	2	US-08-476-082A-45	Sequence 45, Appl
4	108.5	6.6	769	2	US-07-728-235-31	Sequence 31, Appl
5	109.5	6.6	769	4	US-08-938-085A-31	Sequence 31, Appl
6	109.5	6.6	769	5	PCT-US95-04886-1	Sequence 1, Appl
7	109.5	6.6	769	5	PCT-US96-01314-45	Sequence 45, Appl
8	108	6.6	676	3	US-08-630-172-10	Sequence 10, Appl
9	108	6.6	676	4	US-09-375-419-10	Sequence 10, Appl
10	108	6.6	739	4	US-08-444-818-148	Sequence 148, App
11	108	6.6	2995	4	US-08-444-818-138	Sequence 138, App
12	101	6.1	465	2	US-08-833-678A-2	Sequence 2, Appl
13	101	6.1	465	2	US-08-529-169A-2	Sequence 2, Appl
14	101	6.1	590	4	US-08-850-328-3	Sequence 3, Appl
15	101	6.1	631	2	US-08-833-678A-1	Sequence 1, Appl
16	101	6.1	631	4	US-08-529-169A-1	Sequence 1, Appl
17	101	6.1	1021	1	US-07-910-760-12	Sequence 12, Appl
18	101	6.1	1021	1	US-08-440-519-12	Sequence 12, Appl
19	101	6.1	1021	4	US-08-440-549-12	Sequence 12, Appl
20	101	6.1	1786	4	US-08-444-818-54	Sequence 54, Appl
21	101	6.1	2261	4	US-08-444-818-66	Sequence 66, Appl
22	101	6.1	2436	4	US-08-444-818-75	Sequence 75, Appl
23	101	6.1	2772	4	US-08-444-818-89	Sequence 89, Appl
24	101	6.1	2894	2	US-08-466-975A-23	Sequence 23, Appl
25	101	6.1	2894	2	US-08-391-671A-23	Sequence 23, Appl
26	101	6.1	2894	4	US-08-467-902A-23	Sequence 23, Appl
27	101	6.1	2894	4	US-09-275-265-23	Sequence 23, Appl

```
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
FEATURE: /note= "signal sequence"

NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"

NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"

NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"

NAME/KEY: Region
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"

NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"

PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (B2) (from human
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-789-078-1
```

```
Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVTCLKGGTQMCLIDNFTRELKALHLPVQYQDQNNYLOSVPKPMALVGRFLPASTK 60
Db 436 IIVTVQLPQCECRDQSDRSL--CH-----GKGFLECGICRCDTGVIGNCECQTQ 486
QY 61 LNLITQOLEG-----LPSVNGSAFPSTLPGPKITLAGVCDCA----- 103
Db 487 -GRSQELEGSKRDNNISITCSGLGDCVCGCLCHTSYDVP--KLIYGYCECDTNCER 543
QY 104 -SGDFCN-----CNCNCCNHLHDIERF-----RAIKACLGRNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGPGRGLFCGKC--RCHPFGESACOCERTTECL--NPRVCECS-GRGR--- 595
QY 153 VKPOHNKGNCRSGCLKNKYCEYEAQIM-----CSSIC-----KCIQCKNYEESP 198
Db 596 -----CRC-----NVCECHSGVQLPLCQECPCPCSPGKYISCACBCLKFERGP 638
```

```
RESULT 2
US-08-752-633-1
; Sequence 1, Application US/08752633
; Patent No. 5863889
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Stahaan, Teruna
```

```
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
```

0000 / J44 0000

Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGGTQMLCIDSNTRELKALHLVPOYQDQNNYLSQSDVPKPMTALVGRFLPASTK 60
DB 436 IVTVQVLPOQCECRCDQSRDSL--CH-----GKFLGECIGCRDGTGKNCCEQTQ 486
QY 61 LNLITQOQLEGA-----LPSVYNGSAFFSGSTLPGPKITLAGYCDCEFA----- 103
DB 487 -GRSQELEGSCRKDNNSIICSLGDCVCGQCLHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCNN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
DB 544 YNGQVCGGPGRLGFCGKC--RCHPFGSACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKNCRRSCLNKCCEYEAQIM-----CSSIC-----KCIQCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCSPCGKYISCAECLKFEKGP 638

RESULT 5
US-08-938-085A-31
; Sequence 31, Application US/08938085A
; Patent No. 6339148
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quantana, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,085A
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,215
; FILING DATE: 11-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 033070-080210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-085A-31

Query Match 6.6%; Score 109.5; DB 4; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGGTQMLCIDSNTRELKALHLVPOYQDQNNYLSQSDVPKPMTALVGRFLPASTK 60
DB 436 IVTVQVLPOQCECRCDQSRDSL--CH-----GKFLGECIGCRDGTGKNCCEQTQ 486
QY 61 LNLITQOQLEGA-----LPSVYNGSAFFSGSTLPGPKITLAGYCDCEFA----- 103
DB 487 -GRSQELEGSCRKDNNSIICSLGDCVCGQCLHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCNN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
DB 544 YNGQVCGGPGRLGFCGKC--RCHPFGSACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKNCRRSCLNKCCEYEAQIM-----CSSIC-----KCIQCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCSPCGKYISCAECLKFEKGP 638

RESULT 6
PCT-US95-04886-1
; Sequence 1, Application PC/TUS9504886
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbets, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540

OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label- trans
OTHER INFORMATION: /note- "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label- cyto
OTHER INFORMATION: /note- "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
Qy 1 MVIQKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLQSDVPKPMALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRCDQSRSL--CH-----GKFLGCGICRCDTGYIGKNCCEQTQ 486
Qy 61 LNLITQOLEGA-----LPSVNGSAFFSGSTLPGPKITLAGYCDCA----- 103
Db 487 -GRSQLEGGCRDNNIIICGLGDCVGCGLCHTSDVPG--KLIYQYCECDTINCER 543
Qy 104 -SGDFCNN-----CNCNCCNHLHDIERF-----KAIKACLGRLNPEAFQPKIGKQLGN 152
Db 544 YNGQVCGGPGRLCGKGC--RCHPFGESACQERTTEGCL--NPRRVECS-GRGR--- 595
Qy 153 VKPQHKCNCRRSGCLNKYCEYEAQIM-----CSSIC-----KIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQECPCGPGCKYISCAECLKFEKGP 638

RESULT 7
PCT-US96-01314-45
Sequence 45, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaut
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
TITLE OF INVENTION: ANTAGONISTS
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-01314-45
Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
Qy 1 MVIQKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLQSDVPKPMALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRCDQSRSL--CH-----GKFLGCGICRCDTGYIGKNCCEQTQ 486
Qy 61 LNLITQOLEGA-----LPSVNGSAFFSGSTLPGPKITLAGYCDCA----- 103
Db 487 -GRSQLEGGCRDNNIIICGLGDCVGCGLCHTSDVPG--KLIYQYCECDTINCER 543
Qy 104 -SGDFCNN-----CNCNCCNHLHDIERF-----KAIKACLGRLNPEAFQPKIGKQLGN 152
Db 544 YNGQVCGGPGRLCGKGC--RCHPFGESACQERTTEGCL--NPRRVECS-GRGR--- 595
Qy 153 VKPQHKCNCRRSGCLNKYCEYEAQIM-----CSSIC-----KIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGYQLPLCQECPCGPGCKYISCAECLKFEKGP 638
RESULT 8
US-08-630-172-10
Sequence 10, Application US/08630172
Patent No. 6060054
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,172
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-10

Query Match 6.6%; Score 108; DB 3; Length 676;

Best Local Similarity 21.5%; Pred. No. 0.039;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLOSDVPKPMALVGRFLPASTK 60
Db 414 IVTVQVLPOCECRCDQSDRSLSL-CH-----GKFLGCGICRCDTQYIGNKCECQTQ 464
QY 61 LNLITQOLEGA-----LPSVVNGSAFPSTLPGPKKITLAGYCDGCF----- 103
Db 465 -GRSQLESGCRKDNNSIICSLGDCVCGQCLCHTSDVFG--KLIYQYCECDTINCER 521
QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
Db 522 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACOCERTTEGCL--NPRRVECS-GRGR--- 573
QY 153 VKPQHNKGCNRRSGCLKNYCEYEAQIM-----CSSICKICIGCKNYEESP 198
Db 574 -----CRC-----NVCECHSGVQLPLCQPCPCSGKYISCAECLKFEKGP 615

RESULT 9

US-09-375-419-10
Sequence 10, Application US/09375419

Patent No. 6264950

GENERAL INFORMATION:

APPLICANT: Staerz, Uwe

TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T

TITLE OF INVENTION: LYMPHOCYTE VETO

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver

STATE: Colorado

COUNTRY: U.S.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/375,419

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/630,172.

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2879-36

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 676 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-375-419-10

Query Match 6.6%; Score 108; DB 4; Length 676;

Best Local Similarity 21.5%; Pred. No. 0.039;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLOSDVPKPMALVGRFLPASTK 60
Db 414 IVTVQVLPOCECRCDQSDRSLSL-CH-----GKFLGCGICRCDTQYIGNKCECQTQ 464
QY 61 LNLITQOLEGA-----LPSVVNGSAFPSTLPGPKKITLAGYCDGCF----- 103
Db 465 -GRSQLESGCRKDNNSIICSLGDCVCGQCLCHTSDVFG--KLIYQYCECDTINCER 521
QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
Db 522 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACOCERTTEGCL--NPRRVECS-GRGR--- 573
QY 153 VKPQHNKGCNRRSGCLKNYCEYEAQIM-----CSSICKICIGCKNYEESP 198
Db 574 -----CRC-----NVCECHSGVQLPLCQPCPCSGKYISCAECLKFEKGP 615

RESULT 10

US-08-444-818-148

Sequence 148, Application US/08444818

Patent No. 6150087

GENERAL INFORMATION:

APPLICANT: Chien, David Y.

APPLICANT: Rutter, William J.

TITLE OF INVENTION: NANBV Diagnostics and Vaccines

NUMBER OF SEQUENCES: 777

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94508-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,818

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/403,590

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0110.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876

TELEFAX: (508)359-3885

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 739 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-444-818-148

Query Match 6.6%; Score 108; DB 4; Length 739;

Best Local Similarity 24.9%; Pred. No. 0.043;
Matches 56; Conservative 20; Mismatches 79; Indels 70; Gaps 12;

QY 79 SAPP-SGSTLPGPKKITLAGYCDGCFASGDFCNCCNCCNNL-----HHDIERFKA 130
Db 233 SVIPTSQGVVVVATDALMTGY-----TGDFDSVIDCNTCTVQVDFSLDFTFTIETLP 287


```

QY      228 G--LPRFSDRRPSSCISWVVEATCACLLAQAQEAREHCKSL 270
          |         |         |         |         |         |
Db      1580 GENLP-----YLWAYQATVCARAQAAPPSPDQMWKCL 1611
          |         |         |         |         |         |

RESULT 12
US-08-833-678A-2
; Sequence 2, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,678A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/529,169
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-833-678A-2

Query Match           6.1%; Score 101; DB 2; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.11;
Matches 56; Conservative 21; Mismodels 77; Models 72.
```

QY	79	SAFP-SGSTLPGPPKITLAGYCDCAFSGDGFNCNCNCCNL-----HHDIERFKAI	130
Db	232	SVIPTSGDVVVATDALMTGY-----TGDFSDVIDCNTCVTQVDFSLPTFTIETILP	286
QY	131	KACLGRNPEAFQPKIGKQGLG-----NYKPOHNCNCRRSGCLKN--YCCEYCAQIMCSSI	185
Db	287	QDAVSRQTR--RGRTGRGPGIYRFVAPGE-----RPSGMFDSVLCCECYDA-----	331
QY	186	CKCIGCKNYEESPERKTLMSMPNYMOTGGL-----EG-----SHYLPPTKE	226
Db	332	---GCANYELTP-AETTVRLRAYMTNPGLPVQCQDHLERFEGVFTGLTHIDAHFLSQTK	386
QY	227	SG--LPRFSDRRPSSCISMEVVYEATCACLLAQGEAEKEHCKCL	270
Db	387	SGENLP-----YLVAYQATVCAACQAPPPSDMMKCL	419

```

; APPLICANT: SHIRASHI, J.
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS
; TYPE OF INVENTION: C VIRUS INFECTION
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,328
; FILING DATE: 02-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mays, Thomas D
; REGISTRATION NUMBER: 34,524
; REFERENCE/DOCKET NUMBER: 32273-20004.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-850-328-3

```

```

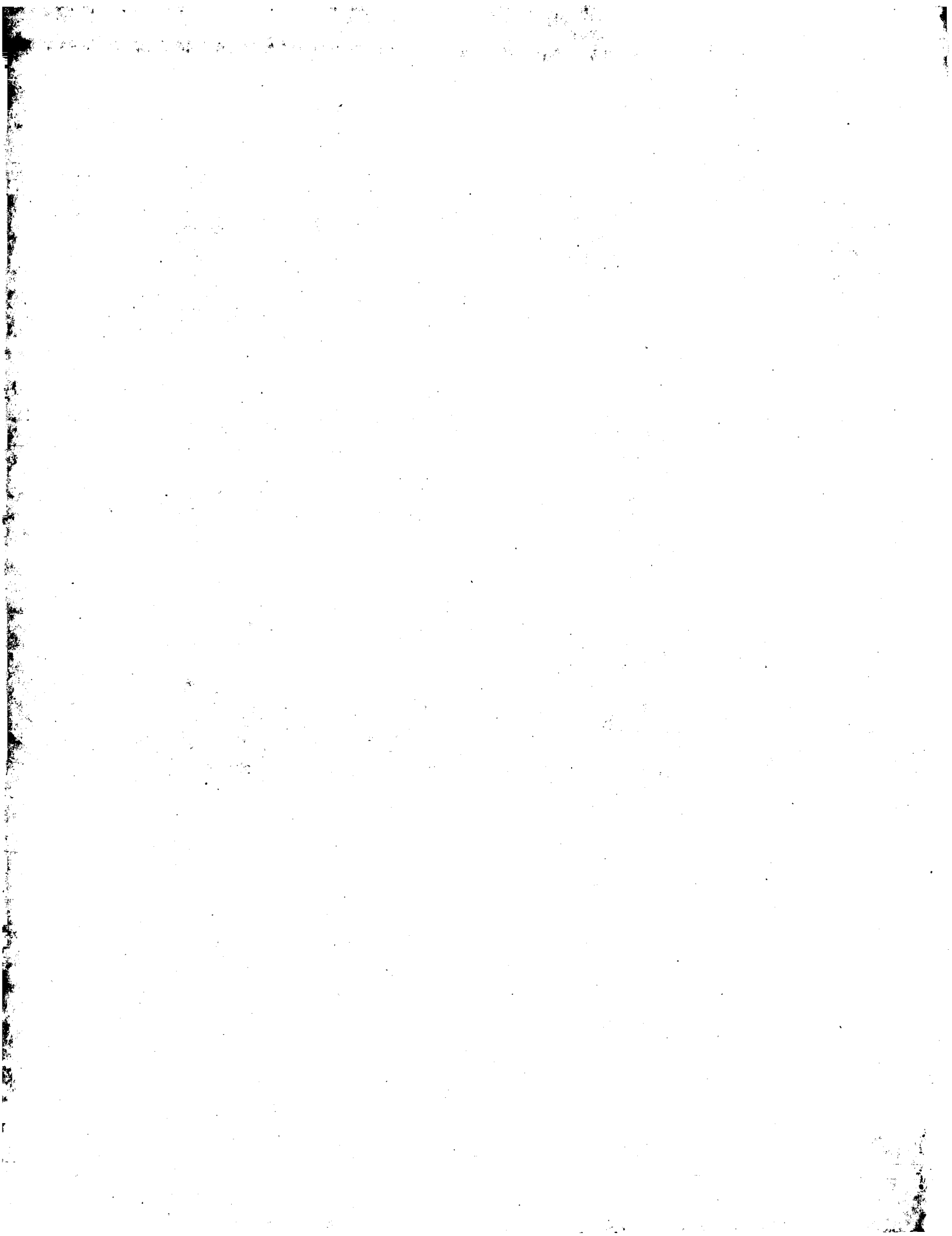
Query Match      6.1%; Score 101; DB 4; Length 590;
Best Local Similarity 24.8%; Pred.No.0.15; Gaps
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps

QY   79 SAFP-SGSTLPGGPKITLAGYCDFASGDGCNNCNCNNCCNL-----HHDIERFKAI 180
Db    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY   131 KACLGRRPEAFQPKIGKGQLG---NVKPHQHKCNCRSGCLKN--YCCEYEAOIMCSSI 185
Db    | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   429 QDAVSRQR--RGRTGRGXPGIYRFVAPGE-----RPSGMFDSSVLCECYDA----- 473
QY   186 CKCIQCKNKYESPERKTLSMPNYMOTGGI-----EG-----SHYLPPTKF 226
Db    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db   474 ----GCAYELTP-AETTVRLRAYMTNPGLPVCODHLEFWEGVTGLTHIDAHFLSQTK 528
QY   227 SG---LPRFSHDRRPSSCISWEVVVEATCALLAQGAEEAHEKSKCL 270
Db    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db   529 SGENLP-----YLWAYQATVCARAQAAPPSPDQMWKCL 561

RESULT 15
US-08-833-678A-1
; Sequence 1, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TYPE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6

```

Search completed: April 21, 2003, 11:51:11
Job time : 15.5909 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:49:09 ; Search time 14.5976 Seconds
(without alignments)
1548.781 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTOMLCIDNSRT.....GRLSQLILHTEFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 75613885 residues
Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437	26.5	438	10	US-09-220-091-13
2	112.5	6.8	1192	9	US-10-189-971-18
3	112.5	6.8	1207	9	US-10-189-971-20
4	112.5	6.8	1477	9	US-10-189-971-8
5	112.5	6.8	1535	9	US-10-189-971-14
6	112.5	6.8	1593	9	US-10-189-971-4
7	109.5	6.6	769	9	US-10-072-841-31
8	106.5	6.5	3635	9	US-10-037-182-4
9	106.5	6.5	3635	10	US-09-845-583-2
10	105.5	6.4	695	10	US-09-864-761-42967
11	105	6.4	1057	9	US-10-189-971-6
12	105	6.4	1251	9	US-10-189-971-16
13	105	6.4	1342	9	US-10-189-971-24
14	105	6.4	1512	9	US-10-189-971-10
15	105	6.4	1570	9	US-10-189-971-12
16	105	6.4	1628	9	US-10-189-971-2
17	103.5	6.3	1725	9	US-10-037-182-12
18	103.5	6.3	1786	9	US-10-037-182-10
19	103.5	6.3	1786	10	US-09-938-275-7

20	102.5	6.2	4123	9	US-10-213-509-5	Sequence 5, Appli
21	101.5	6.2	686	9	US-09-881-239-3	Sequence 3, Appli
22	101.5	6.2	686	10	US-09-881-654-2	Sequence 2, Appli
23	101.5	6.2	728	9	US-09-881-239-1	Sequence 1, Appli
24	101	6.1	1099	10	US-09-881-654-4	Sequence 4, Appli
25	101	6.1	2894	10	US-10-044-995-23	Sequence 23, Appli
26	101	6.1	2894	10	US-09-941-611-23	Sequence 2, Appli
27	101	6.1	3011	10	US-09-916-359-2	Sequence 2, Appli
28	100	6.1	2743	9	US-10-037-182-36	Sequence 36, Appli
29	100	6.1	3695	9	US-10-037-182-2	Sequence 2, Appli
30	99.5	6.0	1551	10	US-09-970-318-4	Sequence 4, Appli
31	98.5	6.0	2749	9	US-10-123-155-385	Sequence 385, App
32	97.5	5.9	632	10	US-09-929-955-29	Sequence 29, Appli
33	97.5	5.9	686	10	US-09-929-955-17	Sequence 17, Appli
34	97.5	5.9	686	10	US-09-929-955-31	Sequence 31, Appli
35	97.5	5.9	686	10	US-09-929-955-32	Sequence 32, Appli
36	97.5	5.9	686	10	US-09-929-955-43	Sequence 43, Appli
37	97.5	5.9	686	10	US-09-929-955-44	Sequence 44, Appli
38	97.5	5.9	686	10	US-09-929-955-45	Sequence 45, Appli
39	97.5	5.9	686	10	US-09-929-955-46	Sequence 46, Appli
40	97.5	5.9	686	10	US-09-929-955-47	Sequence 47, Appli
41	97.5	5.9	686	10	US-09-929-955-48	Sequence 48, Appli
42	97.5	5.9	686	10	US-09-929-955-49	Sequence 49, Appli
43	97.5	5.9	732	9	US-09-977-418-20	Sequence 20, Appli
44	97.5	5.9	1036	9	US-09-373-967-4	Sequence 4, Appli
45	97.5	5.9	1036	9	US-10-028-072-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match	26.5%	Score 437;	DB 10;	Length 438;
Best Local Similarity	41.3%	Pred No. 2	3e-29;	
Matches	92;	Conservative	35;	Mismatches 62; Indels 34; Gaps 8;
QY	98	YDCFPASGDFCNCCNCCNLLHDIERFKAICACLRNPEAFQPKIG--KGOLGNVXP	155	
Db	191	YDCFPANGFECDKCNKCHNNIEYDSQRSLRNPAKPKIGIARGITDIER	250	
QY	156	QHNKGCNRRSCCLKNYCYEAOIMCISICKICGCKN-----YES-PRKTL	203	
Db	251	LHQKCHCKKCKLKNYCYEAKVPTDRCKCKQNTETVMTYKNSGGAVSNAL	310	
QY	204	MSMPNYMQT-----GG-EGSHYLPPTKFSGLPRFSDRR--PSSCSISWEVVEAT	250	
Db	311	MSLTNASSSTATPDGSGSVVTDHGGDY-EDMLLSHKPKVEMDPRFRFPYMTDEVVEAA	369	
QY	251	CACLLAQGEA-----EKEHCCKLAEOIMLEEFGRCLSQL	287	


```

RE-9971-8
US-10-189-971-4
: Sequence 4, Application US/10189971
: Publication No. US20030028907A1
: GENERAL INFORMATION:
: APPLICANT: Walke, D. Wade
: APPLICANT: Scoville, John
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: No. US20030028907A1el Human Kielln-like Proteins and Polynucleoti
: TITLE OF INVENTION: Same
: FILE REFERENCE: LEX-0360-USA
: CURRENT APPLICATION NUMBER: US/10/189, 971
: CURRENT FILING DATE: 2002-07-03
: PRIOR APPLICATION NUMBER: US 60/302, 949
: PRIOR FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: US 60/315, 634

```

```

; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID
US-10-072-841-31

```

Query Match	6.6%	Score 109.5;	DB 9;	Length 769;
Best Local Similarity	22.3%;	Pred. No. 0.26;		
Matches	53;	Conservative 26;	Mismatches 84;	Indels 75; Gaps 15;

QY	1	MVICQLKGGTQMLCIDNSRTRELKALHLVPQYQDNVNVQLSDVPKPMTALVGRFLPASTK	60
		: : : : : : : : : : : : : : : : : : : : : : :	
Db	436	IVTVQVLPQCECRCDQSRDSL--CH-----GKFELEGICRCDTYIGKNCECOTQ	486
QY	61	LNLTQOLEGA-----LPSVVNGSAFPGSTLPGPKKITLAGYDCFA-----	103
		: : : : : : : : : : : : : : : : :	
Db	487	-GRSSQELESGCRKDNNSIIICGIGDCVCGQLCHTSDVPG--KLIYQYCECDTINCR	543
QY	104	-SGDFCNN-----CNCNCCNNLHHDTERF-----KAIKACLAGRNPEAFKIGKGQLGN	152
		: : : : : : : : : : : : : : : :	
Db	544	YNGQVCGGPGRLGLFCGKC--RCHPGFEGSACQERTEGCL--NPRVEGS-GRGR----	595
QY	153	VKPOHNKGCNRRSGCLKNYCEYEAQIM-----CSSITC-----KCIGKNYEESP	198
		: : : : : : : : : : : : : : : :	
Db	596	-----CRC-----NVCEHSYGVLPLQCEBPGCPSPCGYKISCAECLKEKPG	638

```

RESULT 8
US-10-037-182-4
; Sequence 4, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl.
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-P
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-182-4

```

```

Query Match      6.5%; Score 106.5; DB 9; Length 3635;
Best Local Similarity 21.9%; Pred. No. 2.8;
Matches 56; Conservative 26; Mismatches 87; Indels 87; Gaps 15;

QY   94 TLGYCD-CFA-----SGDFCNCCNCCNNLHHDIERKAIKACLGHNPE 139
    ||| ||| | : : : ||| : : : |||
Db   250 TCGGSCDRCPGFNQPWKPATTDSANECOSCNGHAYDCYYDPEVDR-----RNAS 302
    ||| ||| | : : : ||| : : : |||

QY   140 AFQPKIGKGQLGVN---KPOHNGCNRRCGLKNYCCEYAAQTMCSSICKICGNK-- 193
    ||| ||| | : : : ||| : : : |||
Db   303 QNDQNVYQG--GGVCLDCQHHTTGINGER--CLPGFFRAPDQPLSDPHVRPCDCESDFT 358
    ||| ||| | : : : ||| : : : |||

QY   194 ---VEESPERRKTLSMNPYMTGGL-----EG----SHYLPTTKSLGRFSH--DRRPSS 240
    ||| ||| | : : : ||| : : : |||
Db   359 DGTCEDLTGR--CYCRNF--TCELCAACAEGYTDFHCYP-----LPSPFHNDTREQV 408
    ||| ||| | : : : ||| : : : |||

QY   241 CISWEVVVETACLLAGEAEKE-----HCSKC----- 269
    ||| ||| | : : : ||| : : : |||
Db   409 LPAGQIVNCDAAGTQGNACRKDPRLRGVCVKFNFRGAHCELCAPGFHPGSPCHQCQSS 468
    ||| ||| | : : : ||| : : : |||

QY   270 --LAEQMILEEFGRCL 283

```

```

Db      469 PGVANSLCDPESGQCM 484
:| : | |:
RESULT 9
US-09-845-583-2
; Sequence 2, Application US/09845583
; Patent No. US20020142954A1
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert
; APPLICANT: Brunken, William Joseph
; APPLICANT: Champlaud, Marie-France
; APPLICANT: Hunter, Dale
; TITLE OF INVENTION: LAMININ 15 AND USES THEREOF
; FILE REFERENCE: 10287-036001
; CURRENT APPLICATION NUMBER: US/09/845,583
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/200,863
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-845-583-2

```

```

Query Match.          6.5%; Score 106.5; DB 10; Length 3635;
Best Local Similarity 21.9%; Pred. No. 2.8;
Matches 56; Conservative 26; Mismatches 87; Indels 87; Gaps 15;

QY  94  TLACYCD-CFA-----SGDFCNCCNCCNLLHHDIERFKAIKACLGRNPE 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   250  TCGSGCDRCCEGFGNQPPKATPDSANECQSCNCHGHAYDCYYDPEVDR-----RNAS 302

QY  140  AFQPKIGKQGLNV---KPOHKGCCNRRSGCLNKCCEYEAQTMCSSICIGCKN--- 193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   303  QNODNVYQG---GGVCLDCQHHITGINGER---CLPGFFRAPDQPLDSPHVCRPCDSEDT 358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  194  ---YEESPERKTLMSPNYMYOTGGL-----EG-----SHYLPPTKFSGLPRFSH--DRRPSS 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   359  DGTCEDLTGR--CYCRPNF--TGELCAACAEGYTDFPHCYP-----LPSPFHNDTREQV 408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  241  CISWEVEATCACLLAGEAEKE-----HCSKC----- 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   409  LPAGQIVNCDCNAAGTGNCACRKDPRLGRCVKPNFRGAHCCELCAPOGHGSPCHPCOSS 468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  270  --LAEQMILEEFGRCLE 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   469  PGVANSLCDPESGOCM 484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10
US-09-864-761-42967
; Sequence 42967, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USED
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6

```



```

; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo
US-10-189-971-16

```

Query Match 6.4%; Score 105; DB 9; Length 1251;
Best Local Similarity 21.8%; Pred. No. 1.1;
Matches 51; Conservative 19; Mismatches 60; Indels 104; Gaps 13;

[illegible]

```

RESULT 13
US-10-189-971-24
; Sequence 24, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1e1 Human Kielin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302, 949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315, 634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1342
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-24

```

Query Match 6.4%; Score 105; DB 9; Length 1342;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 51; Conservative 19; Mismatches 60; Indels 104; Gaps 13;

QY	77	NGSAFSPGSTLPG--PKTTLACYDCDFASGDFCNN-----CNCNCCNN	119
	:	: :	
Db	556	HGVTTASGETLPDLPTCSL---CTCEGSMRCQKKPCALCPHPSPGPCFPVC---	609
QY	120	LHHDIFFRAIKACIGRNPEAFOPKICKOLGNVKPQHKNCCNRSGCLKNYCCEYEAQ	179
	:	: : : : : :	
Db	610	--HSCLSQGREHODGEFEFGAGSCE-----WCRCQAQG	641
QY	180	IMC-----SSIC-KCIGCKNY-EESPERKTLMSPNTMYQTGLEGSHY	220
	:	: : : : : :	
Db	642	VSCVRQLCPPLCKQLQVTERGSCCPRCGCLAHEHP-----EGSRW	684
QY	221	LPPTKFSGLPRFSDRRRPPSSCIWEVVEATCA---CL--LAOGEEAEKEHCSCK	269
	:	: :	
Db	685	VPP-----DSAGSCSVCEHG-VTCARIOCISSCAPROGHDPCCPOC	726

RESULT 14

```

US-10-189-971-10
; Sequence 10, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielln-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1512
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-10

Query Match 6.4%; Score 105; DB 9; Length 1512;
Best Local Similarity 21.8%; Pred. No. 1.4;
Matches 51; Conservative 19; Mismatches 60; Indels 104; Gaps 13;

QY 77 NGSAPPSSGTLPGP--PKITLAGYCDCFASGDFCNN-----CNENNCCNN 119
Db 726 HGVTTASGETLPDPLDPTCSL---CTCQEGSMRCQKKPCAPALCPHPSPGCFPCVC--- 779
QY 120 LHHDIERFAIKALGRNPFAQFKICKGOLGNVKPOHNGKNCRRSGCLKNYCECYEQ 179
Db 780 -----HSLCSQGRHQDGEFEGPAGSCE-----WCRCOAGQ 811
QY 180 IMC-----SSIC-KCIGCKNY-EESPERKTLMSMPNVMYMTGGLEGSHY 220
Db 812 VSCVRLQCPPLPCKLQVTERGSCCPRCRCGLAHGEEHP-----EGSRW 854
QY 221 LPPTKFSGLPRFSDRRPPSSCIWEVVEATCA---CL--LAQGEAEAEKHCSC 269
Db 855 VPP-----DSACSSCVCHGY-VTCARIQCISSCAQPROGPHDCCPQC 896

RESULT 15
US-10-189-971-12
; Sequence 12, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielln-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1570
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-12

```

Query Match 6.4%; Score 105; DB 9; Length 1570;
Best Local Similarity 21.8%; Pred. No. 1.4;
Matches 51; Conservative 19; Mismatches 60; Indels 10

1	1647	100.0	299	21	US-09-743-237-5	Sequence 5, Appli
2	1351	76.0	295	21	US-09-743-237-4	Sequence 4, Appli
3	1351	76.0	295	21	US-09-743-237-4	Sequence 24, Appli
4	533.5	32.4	223	1	PCR-US01-08656-10693	Sequence 10693, A
5	470	28.5	403	26	US-60-161-932-1700	Sequence 1700, Ap
6	470	28.5	890	26	US-60-167-1217-12948	Sequence 12948, A
7	470	28.5	950	20	US-09-614-150-12897	Sequence 12897, A


```

Db      57 SASPRPCNCTSLKLKLYCDCFANGECFNCCNTNCYNLEHEHNEROKATKACLD RNP    116
QY     139 EAFOPKITGKGQLGNVPQHNGKNCRRRSGCLKNYCCEYEQAICMCSSTCKICGCKNYEESP   198
       :|:||||||:|:||:||||||:||||||:||||||:||||||:||||||:||||||:
Db     117 EAPKPKIGKGESDRRHSKGCNKRRSGLKNYCCEYEAKIMCSSICKICGCKNFEE SP    176
QY     199 ERKTLMSM 206
       ||||||:
Db     177 ERKTLMLHL 184

RESULT          5
US-60-161-932-1700
; Sequence 1700, Application US/60161932
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: CLOO00122
; CURRENT APPLICATION NUMBER: US/60/161,932
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 2626
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1700
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Drosophila
US-60-161-932-1700
```

```

Query Match      28.5%; Score 470; DB 26; Length 403;
Best Local Similarity 32.5%; Pred. No. 1.4e-35;
Matches 110; Conservative 55; Mismatches 97; Indels 76; Gaps 10;

QY      8 GGTOMLC---IDNSRTRELKALHLVPVQODQNNYLQSDV--PKPMTALVGRF-----54
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       65 GAKYMICNSGVPOSTSAMR-----RGTYGYENKTRRPPTISPQOHRFKMGPPQQ 116
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      55 -----LPASTKLNLITQOL-----EGALPSV-----VNGSAFPGSGTLPGPPKIT 94
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       117 SKHQQLQAQAQRIRQQPLPTEQSTPIKVEPKLTPLPGVKANYPAPKPLFEVLKPATAA 176
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY      95 LAG-----YDCGFASGDFCNMCNCCNNCCNLUHDIERFKA 129
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db       177 AAGAVDPLGGMTSRKKHCNCSSKSCLYLKDCFFANGFCQDCTKDCFNNLDYEVRERA 236
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY     130 IKACLRNPFAOPKIKGQGLGNVKPOHNKCNCNRSCRLKNCECYEFAQIMCSICKKI 189
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     237 IRSCLRNPFAFPKPIKTAPNSGDWR-LHNKGCNCKRSGLKNCECYEBAKIPCSICKCV 295
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY     190 GKCYNEPSRKTLMSMPNYMOTGLEGSHTLPPTKFSGLPFRSHDRPPSSCIWSVEVEYA 249
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     296 GCRNEDRPD---VDMSDLGLMGVEQO-----KKDKRAKNQLNENRANIYFTDDVIEA 346
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY     250 TCACLLAQGBAEKEHKSCKLAEOIMILEEFGRCLISQIL 287
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db     347 TIMCMISRIVHMHEKONAVEMEREVMEGMESLTQII 384
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

```

RESULT      6
US-60-167-217-12948
US-60-167-217, Application US/60167217
; Sequence 12948, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSEQ for Windows Version 4.0

```

```
; SEQ ID NO 12948
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-12948

Query Match      28.5%   Score 470; DB 26; Length 890;
Best Local Similarity 32.5%; Pred. No. 4.1e-35;
Matches 110; Conservative 55; Mismatches 97; Indels 76; Gaps 10;
```

```

RESULT 7
US-09-614-150-12897
; Sequence 12897, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12897
; LENGTH: 950
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150-12897

```

[illegible]


```

RESULT 13
US-09-087-136-13
; Sequence 13, Application US/09087136
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

Query Match          26.5%; Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

QY 98 YCDCFASGDFCNCCNCCNNLHHDIERKAKACALGRNPEAFOPKIG--KGOLGNVKP 155
DB 191 YCDCFANGFCRCDCNCKDCHNNIYDSQRSKALRQSLERNPNAPKPKIGIARGITDIER 250
QY 156 QHNKGCMRRSGCLKNYCEYEAQIMCSSICKICGCKN-----YEES-----PERKTL 203
DB 251 LHQKGCHCKSGCLKNYCEYEAQVPCPTRCKCKGCONETRYMTRYKNSGGAVSNTNAL 310
QY 204 MSPNVMQT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISWEVVVEAT 250
DB 311 MSLTNASSTATPDSPGGSVVYTDHGGDY--EDMLLSHKPKVEMDPRRPWYMTDEVVEAA 369
QY 251 CACLLAAGEA-----EKHCSCKLAEQMLEEFGCLSQL 287
DB 370 TMCVAAQEALNVEKQVTEDEKLINMEKLVLEFGCLQMI 412

RESULT 14
US-09-087-136-13
; Sequence 13, Application US/09087136A
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136A
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

Query Match          26.5%; Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

QY 98 YCDCFASGDFCNCCNCCNNLHHDIERKAKACALGRNPEAFOPKIG--KGOLGNVKP 155
DB 191 YCDCFANGFCRCDCNCKDCHNNIYDSQRSKALRQSLERNPNAPKPKIGIARGITDIER 250
QY 156 QHNKGCMRRSGCLKNYCEYEAQIMCSSICKICGCKN-----YEES-----PERKTL 203
DB 251 LHQKGCHCKSGCLKNYCEYEAQVPCPTRCKCKGCONETRYMTRYKNSGGAVSNTNAL 310
QY 204 MSPNVMQT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISWEVVVEAT 250
DB 311 MSLTNASSTATPDSPGGSVVYTDHGGDY--EDMLLSHKPKVEMDPRRPWYMTDEVVEAA 369
QY 251 CACLLAAGEA-----EKHCSCKLAEQMLEEFGCLSQL 287
DB 370 TMCVAAQEALNVEKQVTEDEKLINMEKLVLEFGCLQMI 412

```


Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	366.5	22.3	282	6	US-10-155-881-9626	Sequence 9626, Ap
2	357	21.7	603	5	US-09-573-6558-2154	Sequence 2154, Ap
3	357	21.7	603	5	US-09-573-6558-2192	Sequence 2192, Ap
4	348	21.1	510	6	US-10-155-881-9875	Sequence 9875, Ap
5	278	16.9	53	6	US-10-106-698-6932	Sequence 6932, Ap
6	248.5	15.1	896	6	US-10-155-881-20031	Sequence 20031, A
7	201.5	12.2	359	6	US-10-155-881-9183	Sequence 9183, Ap
8	183	11.1	143	6	US-10-155-881-9877	Sequence 9877, Ap
9	167	10.1	120	6	US-10-138-145-1127	Sequence 1127, Ap
10	112	6.8	1774	1	PCR-US02-01339-6	Sequence 6, Appli
11	109.5	6.6	599	1	PCR-US02-09671-718	Sequence 718, App
12	109.5	6.6	761	1	PCR-US02-09671-711	Sequence 711, App
13	109.5	6.6	769	1	PCR-US02-09671-708	Sequence 708, App
14	109.5	6.6	769	1	PCR-US02-09671-710	Sequence 710, App
15	109.5	6.6	769	1	PCR-US02-09671-712	Sequence 712, App
16	109.5	6.6	769	1	PCR-US02-09671-713	Sequence 713, App
17	109.5	6.6	769	1	PCR-US02-09671-714	Sequence 714, App
18	109.5	6.6	769	1	PCR-US02-09671-715	Sequence 715, App
19	109.5	6.6	769	1	PCR-US02-09671-716	Sequence 716, App
20	109.5	6.6	769	1	PCR-US02-09671-719	Sequence 719, App
21	109.5	6.6	769	1	PCR-US02-09671-720	Sequence 720, App
22	109.5	6.6	769	6	US-10-072-844-31	Sequence 31, Appl
23	109.5	6.6	769	6	US-10-072-838-31	Sequence 31, Appl
24	109.5	6.6	769	6	US-10-099-007A-12	Sequence 12, Appl
25	109.5	6.6	817	6	US-10-144-793-141	Sequence 141, App
26	108	6.6	576	1	PCR-US02-09671-717	Sequence 717, App

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match 16.9%; Score 278; DB 6; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.3e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 VEATCACLAAQGEAEKEKCSKLAQEMILEEFGRCLSQLILTEPKSKGLKME 299
Db 1 VEATCACLAAQGEAEKEKCSKLAQEMILEEFGRCLSQLILTEPKSKGLKME 53

RESULT 6
US-10-155-881-20031
; Sequence 20031, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 20031
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-20031

Query Match 15.1%; Score 248.5; DB 6; Length 896;
Best Local Similarity 40.7%; Pred. No. 5.3e-13;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;
QY 98 YDCCFASGDFCNCNCCNHLHDIERFKAIKACL-GRNPEAFQPKIGKQGLGNVXP 155
Db 487 YDCCFAGTYCTDPCACQGLNRPEY-VETVYETKQIESRNPFIAPAKI-----VQP 538
QY 156 -----QHNGKNCRRSGCLKNKYCEYEAQIMCSSICKIGCKNYEE 196
Db 539 TTDISSHMDENLTPSSARHKGNCNCRSMCLKKYCEYQANVGSSGCRCEGCKNVHG 598
QY 197 SPE 199
Db 599 KKE 601

RESULT 7
US-10-155-881-9183
; Sequence 9183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9183
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9183

Query Match 12.2%; Score 201.5; DB 6; Length 359;
Best Local Similarity 23.1%; Pred. No. 1.8e-09;
Matches 63; Conservative 35; Mismatches 102; Indels 73; Gaps 9;
QY 98 YDCCFASGDFCNCNCCN-----NLHHD-----IERFKAIKACLGRNPEAFQPKIG 146
Db 80 YPCFSGGGYCDKCGCQPCFNKEAFETVHTTRKVLLSRQKMSMKINRPEANTEPME 139
QY 147 KQQLGNVQPHNKGNCRRSGCLKNKYCEYEAQIMCSSICKIGCKNYEESPERKTLMSM 206
Db 140 DAHSSSSSTPPKRGKNCCKSSCLKKYCDYQDGTGCSLFCRCDDCONPFGKNEGIMADDS 199
QY 207 PNYMOTGG-----LEGS-----HYLPP-----TK 225
Db 200 KRYLTGADLDHSEGEHDFVVERSPLOSPIKSESFHOTPPHLRASSRDAHVFPHAISQ 259
QY 226 FSLGPRFSH--DRPSSCIS-----WEVVE-----ATCACLAAQGEAEKE 264
Db 260 WQALPRSMHCSNKRNSDRAMDSDANYKNSHDYQVQPKHEDSYSISKVCQILNGMAELS 319
QY 265 HCSKCLAEQMLEEFGRLCSQLILTEFKSKGLK 297
Db 320 QVEKSVAPDVFQPGNREIFVSLSGDVRAMWLK 352

RESULT 8
US-10-155-881-9877
; Sequence 9877, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9877
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9877

Query Match 11.1%; Score 183; DB 6; Length 143;
Best Local Similarity 40.9%; Pred. No. 2.4e-08;
Matches 36; Conservative 16; Mismatches 24; Indels 12; Gaps 3;
QY 98 YDCCFASGDFCNCNCCNHLHDIERFKAIKACLGRNPEAFQPKIGK-GQLGNV---- 153
Db 55 YCECFQELQYCDGCGNCNCGNIVGNENARNEAIEAIRQNPFAQPKIGNDSNTLNVKRD 114
QY 154 -----KPOHNKGNCRRSGCLKNYCE 174
Db 115 KFWSSPSRSKPHKCHCKKS-CLKKYCD 141

RESULT 9
US-10-138-145-1127
; Sequence 1127, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen

APPLICANT: Shenk, Michael Andrew
APPLICANT: Gibson, John Bryan
APPLICANT: Norriss, Michael Geoffrey
APPLICANT: Saulsbury, Keith Martin
APPLICANT: Hall, Claire
TITLE OF INVENTION: Compositions isolated from forage
TITLE OF INVENTION: grasses and methods for their use.
FILE REFERENCE: 11000.10580
CURRENT APPLICATION NUMBER: US/10/138,145
CURRENT FILING DATE: 2002-05-01
NUMBER OF SEQ ID NOS: 1552
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1127
LENGTH: 120
TYPE: PRT
ORGANISM: Lolium perenne
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(120)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 10.1%; Score 167; DB 6; Length 120;
Best Local Similarity 37.4%; Pred. No. 4.4e-07;
Matches 40; Conservative 15; Mismatches 36; Indels 16; Gaps 8;
QY 98 YDCDFASGDFCNN-CNCCNCLNHHDIERKAKACGLGRNPEAFQPKI-----GKGOLGN 152
Db 16 YCEC-AAGVYCEPCSCGGLNPHIEIV-LSTRKQIEFRNPLA-APKVRLSDAQAQETQ 72
QY 153 VKP-----QHNKGCNRRSGCKLNKCYEYEAQIMCSCICKICGCKN 193
Db 73 EDNTPASARHKRCNCKKSSCLK-YCECYGGVGLTNCRC-ECKN 117

RESULT 10
PCT-US02-01339-6
Sequence 6, Application PC/TUS0201339
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: LEE, Ernestine A.
APPLICANT: WALIA, Narinder K.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: YUE, Henry
APPLICANT: THANGAVELU, Kavitha
APPLICANT: XU, Yuming
APPLICANT: ARVIZU, Chandra
APPLICANT: WARREN, Bridget A.
APPLICANT: YAO, Monique G.
APPLICANT: AU-YOUNG, Janice
APPLICANT: HAFALIA, April J.A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: KALLICK, Deborah A.
APPLICANT: GANDHI, Ameena R.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: KHAN, Farrah A.
APPLICANT: LU, Yan
APPLICANT: SWARNAKAR, Anita
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: NGUYEN, Dannel B.
APPLICANT: GRAUL, Richard
APPLICANT: LU, Dyung Aina M.
TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PI-0346 PCT
CURRENT APPLICATION NUMBER: PCT/US02/01339
CURRENT FILING DATE: 2002-01-16
PRIOR APPLICATION NUMBER: 60/262,838; 60/265,927; 60/271,196; 60/274,549; 60/334,179
PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PERL Program

SEQ ID NO 6
LENGTH: 1774
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7722591CDI
PCT-US02-01339-6

Query Match 6.8%; Score 112; DB 1; Length 1774;
Best Local Similarity 20.2%; Pred. No. 0.46;
Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;
QY 86 TLPQPKITLAGYCDGFCFNN---CN-CNNCC---NNLH-----H 122
Db 800 TLPDP-----CRLLSSPEACNQSACTWCHGACLSGDQAHRLCGGSPSPMPRSPE 852
QY 123 DIERFKAIKACGLGRNPEAFQPKIGKQGLGNVKNPKQHNKNCRRSGCL-----KNYCE 174
Db 853 ECRRLRTCSCLARHPRTLQP--GDGEAST--PRCKWCTNCPGACIGRNGSCTSENDNR 908
QY 175 CYEAQIMCSCIC-----KICCKNYEESPERKTLMSM-PNY----- 209
Db 909 INOREVFWAGNCSEAAAGAACDQCCTREGKCMWTRQFKRTGETRRLSVQPTYDWTCTFSH 968
QY 210 ----MOTGGLGSHYLP-PTKFSGLPRFSDHRRPSSCISWEVVEATCACLAGEAEAEKE 264
Db 969 SLNVSPMPVESFPPLPCTPPCHLLP-----NCTS-----CLDSKAGDGGWQ 1010
QY 265 HC-----SKCLA 271
Db 1011 HCVWSSSLQOCLS 1023

RESULT 11
PCT-US02-09671-718
Sequence 718, Application PC/TUS0209671
GENERAL INFORMATION:
APPLICANT: ZYCOs Inc.
TITLE OF INVENTION: TRANSLATIONAL PROFILING
FILE REFERENCE: 08191-026WOI
CURRENT APPLICATION NUMBER: PCT/US02/09671
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: 60/279,495
PRIOR FILING DATE: 2001-03-28
PRIOR APPLICATION NUMBER: 60/292,544
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: 60/310,801
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/326,370
PRIOR FILING DATE: 2001-10-01
PRIOR APPLICATION NUMBER: 60/336,780
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 60/358,985
PRIOR FILING DATE: 2002-02-20
NUMBER OF SEQ ID NOS: 2041
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 718
LENGTH: 699
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-09671-718

Query Match 6.6%; Score 109.5; DB 1; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVTCOLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOCECRCDQSRDRSL--CH-----GKGFLECGICRCDTYIGNKCECQTQ 486

Job time: 40 sec

Best Local Similarity 22.3%; Pred. No. 0.29;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

1	QY	MYTCQLKGTOMLCTDNSTREKALHLVPQYQDQNNYLQSDVPKPMVALVGRFLPASTK	60
436	Db	IVTVQVLPQCECRCDQSDRSJL---CH-----GKGFLECGICICDGTGYIGKNECQTQ	486
61	QY	LNLLITQOLEGA-----LPSVYNGSAFPSSGSLPGPPKTLIAGYCDCA-----	103
487	Db	GRSSQLEGGSCRKDNNSIICSLGDCVCGCQLCHTSDVPG---KLIIYGYCECDTINCR	543
104	QY	SGDFCNN-----CNCNCCNLLHHDIERF-----KAIKACLGRNPEAFQPKIGKGLGN	152
544	Db	YNGQVCGGPGRLGFCGKC--RCHPFGESACOCERTTEGCL--NPRRVBCS-GRGR---	595
153	QY	VRPQHNGKNCRRSGLKKNYCYEQAQIM-----CSSIC-----KCI GKKNYEEPS	198
596	Db	-----CRC-----NVCECHSGYUPLQCECGPGGSPCGKYISCAECLKFKSGP	638

```

RESULT 15
PCT-US02-09671-712
; Sequence 712, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: zycos Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026W01
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 712
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-712

```

Query Match	6.6%	Score 109.5	DB 1	Length 769
Best Local Similarity	22.3%	Pred. No. 0.29		
Matches	53	Conservative 26	Mismatches 84	Indels 75
				Gaps 15

QY	1	MVTCQLKGGTQMLCINSTRRELKALHLVPOYQDNNYNLOSDFPKPMATLVGRLPASTK	60
DB			
QY	436	IVTVQVLPQCECRQDSRDSL-CH-----GKFLGCGTCRCDTGYIGNKCECQTQ	486
DB			
QY	61	LNLIITQOLEG-----LPSVVNGSAPFSGTGLPGPKKITLAGYDCFA-----	103
DB			
QY	487	-GRSSQLEGGSCRRDNNISICSGLGDCVCGQCILCHTSDVPG--KLIYGYQCECDTINCER	543
DB			
QY	104	-SGDFCNN-----CNCNCCNNLHHDIERE-----KAIKACLGRNPEAFQKIGKQGLGN	152
DB			
QY	544	YNGQVCGGPGRLGFCGKC--RCHPGFEGSACQERTEGCL--NPRVCS-GRGR---	595
DB			
QY	153	VRPQHNGKNCRSGCLKNVCYCEYEAQIM-----CSSIC-----KCIGCKNVFEESP	198
DB			
QY	596	-----CR-----NVCECHSGYQLPQCECPGCPSPCKGYISCAELCKPEKG	638
DB			

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:45:13 ; Search time 15.6044 Seconds
(without alignments)
1842.060 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTEFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	26.5	429	2 T23152	hypothetical prote
2	376	22.8	571	2 B84585	hypothetical prote
3	357	21.7	603	2 T08955	hypothetical prote
4	245.5	14.9	658	2 F71410	hypothetical prote
5	117	7.1	497	2 T27827	hypothetical prote
6	112	6.8	1737	2 T00209	MEGF8 protein - hu
7	109.5	6.6	769	1 L1JHLM	leukocyte adhesion
8	106.5	6.5	3635	2 T10053	laminin alpha 5 ch
9	105.5	6.4	2180	2 T29784	hypothetical prote
10	104	6.3	775	2 T21436	hypothetical prote
11	103.5	6.3	1786	1 MMSB1	laminin beta-1 cha
12	102	6.2	1790	1 MNFEB1	laminin beta-1 cha
13	101	6.1	1700	2 S08167	Balbani ring 3 pr
14	101	6.1	1819	2 A71928	cag island protein
15	101	6.1	3011	1 GNWVC3	genome polyprotein
16	100.5	6.1	633	2 T24898	hypothetical prote
17	100	6.1	395	2 S4454	transcription fact
18	100	6.1	520	2 G84510	probable receptor-
19	99.5	6.0	677	2 C42125	trophozoite cystei
20	99	6.0	373	2 T47488	hypothetical prote
21	99	6.0	565	2 T16408	hypothetical prote
22	99	6.0	2195	2 T34264	hypothetical prote
23	98.5	6.0	574	2 B88465	protein B0244:8 [i
24	98.5	6.0	962	2 JC5571	subtilisin-like pr
25	98.5	6.0	3106	1 S53868	laminin alpha-2 ch
26	98.5	6.0	5376	2 T42215	zonadhesin - mouse
27	97.5	5.9	432	2 T37509	hypothetical prote
28	97.5	5.9	969	1 A39490	subtilisin-like pr
29	97.5	5.9	975	2 JC5570	subtilisin-like pr

30	97.5	5.9	1766	2 A42125	trophozoite cystei
31	97	5.9	382	2 T29339	hypothetical prote
32	97	5.9	798	2 A40526	integrin beta-7 ch
33	97	5.9	3010	1 A40573	genome polyprotein
34	96.5	5.9	290	2 G72858	AcOrf-70 protein -
35	96.5	5.9	1895	2 T15881	hypothetical prote
36	96.5	5.9	3759	2 A35085	trithorax protein
37	96	5.8	798	2 A28193	integrin beta-1 ch
38	96	5.8	3010	1 GNWVC	genome polyprotein
39	95.5	5.8	1801	1 MNRTS	laminin beta-2 cha
40	95	5.8	1927	2 G64585	cag pathogenicity
41	94	5.7	3011	1 GNWVC	genome polyprotein
42	93.5	5.7	772	2 S32659	integrin beta 2 ch
43	93.5	5.7	1599	2 T16210	hypothetical prote
44	93	5.6	1751	1 MMHUMH	laminin alpha-2 ch
45	93	5.6	1810	1 A32230	tenascin precursor

ALIGNMENTS

RESULT 1

T23152
hypothetical protein JC8.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23152
R:Lightning, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19699
A:Accession: T23152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-429 <WIL>
A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b
A:Experimental source: clone JC8
C:Genetics:
A:Gene: CESP:JC8.6b
A:Map position: 4
A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match	26.5%	Score 437;	DB 2;	Length 429;
Best Local Similarity	41.3%	Pred. No. 6.9e-28;		
Matches	92;	Conservative	35;	Mismatches 62;
				Indels 34;
				Gaps 8;
Qy	98	YDCDFASGDFCNCCNCCNHHDIERFKAKACLRNPEAFQPKIG--KGQLGNVYP	155	
Db	182	YDCDFANGFCRDCNCKDCHNNIEYDSQSKAIRSLERNPNAPFKIGIARGGITDIER	241	
Qy	156	QHNKGCNRRSGCLKNYCEYEAQIMCSSICKICCKN-----YEES---PERKTL	203	
Db	242	LHQGCHCKKSGCLKNYCEYEAQVPTDRCKCKCQCTYRMTRYKNSGAVSNTNAL	301	
Qy	204	MSMPNYMT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISNWEAT	250	
Db	302	MSLTNASSTATPDSGFSVVTDEHGGDY-EDMLLSHKPKVEMDPRFRFYMTDEVAA	360	
Qy	251	CACLLAQGEA-----EKEHCKSLAEQIMLEEFGRCLSQL	287	
Db	361	TMCNVAQAEEALNKEVKQTEDEKLNMEKLVLEFRGRCLEQMI	403	

RESULT 2

B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Experimental source: clone ZK287

C:Genetics:

A:Gene: CESP:ZK287.1

A:Map position: 5

A:Introns: 20/1: 44/2; 99/3; 339/3; 443/2

Query Match 7.1%; Score 117; DB 2; Length 497;

Best Local Similarity 23.3%; Pred. No. 0.067;

Matches 55; Conservative 26; Mismatches 81; Indels 74; Gaps 13;

QY 99 CDCFASGDFCENN-CNCNCCNLLHDIERFKAICACLRNPEAFQPKICKGOLGNV-KPQ 156

DB 77 CDCRLKECKSLCAKQIC-----KNKEA--PK-----KLAKYAKP- 111

QY 157 HNKGNCR--RSGLKNYCEYEAQIMCSSICKIG-CNKYEESPERKTLMSNPYMQG 213

DB 112 -TSGCOCAGKKQCVKKEACRTVYGFCSACKCGGDCYN-----GASKFSVPKHVQNC 164

QY 214 GLEGSHYLPPTKPSGL-----PRFSDRRPSSCISWEVVEATCACL----- 254

DB 165 FLEKH-----ESSGLIVTLIGEDVYRGDFYHESKGEHPVEEQLVAIYDLISKYVDL 219

QY 255 -----LAQGEAEKEHCKSLAEQIMLEBFGRCLSQILHTEPKSKGLKM 298

DB 220 HEIQIFVSKSPCFHODCEPKCEVDECKSNKACAKILLGLLSKV-RKEIKKVDVKM 274

RESULT 6

T00209

MEGF8 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00209

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs

A:Reference number: 214126; MUID:98360089; PMID:9693030

A:Accession: T00209

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1737 <NAK>

A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308

A:Experimental source: brain; clone HGI392

C:Genetics:

A:Gene: MEGF8

A:Map position: 19q12

Query Match 6.8%; Score 112; DB 2; Length 1737;

Best Local Similarity 20.2%; Pred. No. 0.57;

Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;

QY 86 TLPGPPKITLAGYCDCFASGDFCENN---CN-CNNCC---NNLH-----H 122

DB 763 TLPPDP-----CRLLSSPEACNSGACTWCHGACLSGDQAHRLGCGSPGSPRSP 815

QY 123 DIERFKAICACLRNPEAFQPKIGKLGOLGNVQPNHKGNCNRRSGCL-----KNYCE 174

DB 816 ECRRLTSCCLARHPTLQP--GDGEAST--PRCKWCTNCPGACIGRNGSCTSENC 871

QY 175 CYEAQIMCSSIC-----KCIGCKNYEESPERKTLMSM-PNY----- 209

DB 872 INQREVFWAGNCSAAGAACDCEQCTREGKCMWTQFKRTGTETRLISVOPTYDTCFSH 931

QY 210 ----NOTGLEGSHYLP-PTKESGLPRFSDRRPSSCISWEVVEATCACLLAQGEAEKE 264

DB 932 SLNVSMPVSSPPPLPCPTPCHLP-----NCTS-----CLDSKGADGGWQ 973

QY 265 HC-----SKCLA 271

DB 974 HCWSSSLQOCLS 986

RESULT 7

IJHULM

leukocyte adhesion protein beta chain (CD18) precursor - human

N:Alternate names: integrin beta-2; LFA-1 beta chain; Mac-1 beta chain; p150,95 beta

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999

C:Accession: A25967; A91084; S19324; I70090; I70091; A29265; A94497

R:Kishimoto, T.K.; O'Connor, K.; Lee, A.; Roberts, T.M.; Springer, T.A.

Cell 48, 681-690, 1987

A:Title: Cloning of the beta subunit of the leukocyte adhesion proteins: homology to

A:Reference number: A25967; MUID:87131080; PMID:3028646

A:Accession: A25967

A:Molecule type: mRNA

A:Residues: 1-769 <KIS>

A:Cross-references: GB:M15395; NID:q186933; PIDN:AAA59490.1; PID:g307113

A:Note: source of LFA-1 was the SKW3 T-cell line; source of Mac-1 was pooled leukocyte

R:Law, S.K.A.; Gagnon, J.; Hildreth, J.E.; Wells, C.E.; Willis, A.C.; Wong, A.J.

EMBO J. 6, 915-919, 1987

A:Title: The primary structure of the B-subunit of the cell surface adhesion glycopro

A:Reference number: A91084; MUID:87246525; PMID:2954816

A:Accession: A91084

A:Molecule type: mRNA

A:Residues: 9-198, 'P', 200-769 <LA2>

A:Cross-references: GB:Y00057; NID:g30228; PIDN:CAA68266.1; PID:g7629339

R:Wetzman, J.B.; Wells, C.E.; Wright, A.H.; Clark, P.A.; Law, S.K.A.

FEBS Lett. 294, 97-103, 1991

A:Title: The gene organisation of the human beta2 integrin subunit (CD18).

A:Reference number: S19324; MUID:92077153; PMID:1683838

A:Accession: S19324

A:Molecule type: DNA

A:Residues: 1-3;16-23;46-53;106-113;163-170;244-251;296-303;328-335;358-365;405-412;4

R:Nelson, C.; Rabb, H.; Arnaout, M.A.

J. Biol. Chem. 267, 3351-3357, 1992

A:Title: Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing a

A:Reference number: I35376; MUID:92147694; PMID:1346613

A:Accession: I70090

A>Status: translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 347-350, 'S', 352-355 <NEL>

A:Cross-references: GB:S81234; NID:g245294; PIDN:AAB21404.1; PID:g245295

A:Note: mutant form

A:Accession: I70091

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 246-247, 'PSSO', 248-249 <RES>

A:Cross-references: GB:S81252; NID:g245289; PIDN:AAB21402.1; PID:g245290

A:Note: mutant splice form

C:Comment: The leukocyte adhesion proteins are noncovalently linked heterodimers of d

and a deficiency of them is attributed to a genetic defect in the expression or stru

C:Comment: The cysteine residues are involved in intrachain disulfide bonds.

C:Genetics:

A:Gene: GDB:ITGB2

A:Cross-references: GDB:120574; OMIM:600065

A:Map position: 21q22.3-21q22.3

C:Superfamily: integrin beta chain; laminin-type EGF-like homology

C:Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; leuk

F:23-769/Product: leukocyte adhesion protein beta chain #status predicted <NA>

F:23-670/Domain: extracellular #status predicted <EXT>

F:445-631/Region: cysteine-rich

F:459-540/Region: duplication

F:541-627/Region: duplication

F:701-723/Domain: transmembrane #status predicted <MEM>

F:724-769/Domain: intracellular #status predicted <CYT>

F:23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic

F:50;116,212,254,501/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.6%; Score 109.5; DB 1; Length 769;

Best Local Similarity 22.3%; Pred. No. 0.41;

Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MYICQLKGGTQMLCIDSNTRELKALHLVQYQDONNYLQSDVPKMTALVGRFLPASTK 60

DB 436 IYTVQVLPOCECRCDQSRDSL--CH-----GKGFLECGICRCDTGYIGNCBQCNQ 486

[illegible]

A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PID:AA06047.1; PID:g415500
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 6.1%; Score 101; DB 2; Length 1819;

Best Local Similarity 22.7%; Pred. No. 4.7;

Matches 75; Conservative 40; Mismatches 105; Indels 110; Gaps 19;

QY 14 CINDSTRE--LKALHLVQYQDNNYLOSVDVPRKMTALVGRPLPASTKLNLTQQ--LEG 70

Db 363 CIKNAETEERIKCLDLI--KDN-----LKKSLNNOQKQVQ 397

QY 71 ALPSVYNGSAFFSGS---TLPGPKI-----TLAGYCDCAFSGDFCNCCNN 115

Db 398 ALDCLANAKTDERKECKLINDPEYREKRELEQKLEQKDYKDKIKNAK--TEAEKNE 455

QY 116 CCNNLHHD--IERPK-----AIKACLRNPEAFQPKIGKQGLGNVYKPOHNKG 160

Db 456 CLKGLSKEAIERLKQALDCLNNAKTDERKECKLNIPQDLQEL----LADMSVKAYKD 511

QY 161 CNCR--RSGCLKNYCE--CYEAQIMCSSICKICGC--KNYEESPERK--TLMSMPNYWOTGG 214

Db 512 CVSRARNEKEKQCEKLLTPEAKLLEN--QALDCLNNAKTDERKECKLNLPKDLQ--- 566

QY 215 LEGSHYLPPTKFSGLPRFSDRRPSSCISWEVVEATCACLLAQGEAEKEHCKSKLAEQ- 273

Db 567 -----SDILAKESLKAYKDCASQAKTEAEKKECKLLTPEA 602

QY 274 -MILEEFG-----CLSQILHTEFKSK 294

Db 603 KKLLEEEAKESVKRAYLDCVVSQ-AKTEAEKK 631

RESULT 15

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Co

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826; PMID:1848704

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PID:AAA45676.1; PID:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.H

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e

A:Reference number: PQ0393; MUID:92268871; PMID:1316939

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DDBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepatitis C virus #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 6.1%; Score 101; DB 1; Length 3011;

Best Local Similarity 24.8%; Pred. No. 7.6;

Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAPP--SGSTLPGPPKITLAGYCDCAFSGDFCNCCNCCNNL-----HHDIERFKAI 130

Db 1424 SVIPTSDDVVVATDALMTGY-----TGDFSDVDCNTCTVQTFDFSLDPTFTIETILP 1478

QY 131 KACLRNPEAFQPKIGKQGLG---NVKQHNKCCNRRSGCLKN--YCCEYEAQIMCSSI 185

Db 1479 QDAVSRTQR--RGRTGRGRPGIYRFVAPGE-----RPSGMFDSVLCCEYDA----- 1523

QY 186 CKICGCKNYEESPERKTLMSMPNYMOTGGL-----EG-----SHVLPPTKF 236

Db 1524 ----GCANYELTP-AETTVRLRAYMNTPLGVQCQDHLFEWEGVFTGLTHDAHFLSQTKQ 1578

QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLLAQGEAEKEHCKSKCL 270

Db 1579 SGENLP-----YLVAYQATVCARQAQAPPSPMDQMNKCL 1611

Search completed: April 21, 2003, 11:50:45

Job time : 21.6044 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:43:43 ; Search time 28.1886 Seconds
(without alignments)
2185.571 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRLSLILHTEFKSKGLKWE 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1251	76.0	475	11 Q8VIE0	Q8vie0 mus musculu
2	1244	75.5	475	11 Q8VIE1	Q8vie1 mus musculu
3	470	28.5	950	5 Q9V6Q8	Q9v6q8 drosophila
4	437	26.5	429	5 O62295	O62295 caenorhabdi
5	437	26.5	435	5 Q95QD7	Q95qd7 caenorhabdi
6	376	22.8	571	10 Q9SL70	Q9sl70 arabidopsis
7	369	22.4	277	11 Q9D571	Q9d571 mus musculu
8	357	21.7	603	10 Q9SZD1	Q9szd1 arabidopsis
9	262.5	15.9	356	10 Q9LW71	Q9lw71 arabidopsis
10	261	15.8	601	10 Q9CAV1	Q9cav1 arabidopsis
11	248.5	15.1	896	10 Q9ZS22	Q9zs22 glycine max
12	247.5	15.0	609	10 Q9M679	Q9m679 arabidopsis
13	245.5	14.9	658	10 Q23333	Q23333 arabidopsis
14	229.5	13.9	526	10 Q94A12	Q94a12 arabidopsis
15	229.5	13.9	695	10 Q9LE32	Q9le32 arabidopsis
16	229.5	13.9	695	10 Q9LUI3	Q9lui3 arabidopsis

```

17 202.5 12.3 593 10 Q9LUI5
18 200 12.1 243 5 Q9VMO3
19 187.5 11.4 553 10 Q94DS2
20 117 7.1 497 5 Q23460
21 113 6.9 1981 5 Q9VVK7
22 112 6.8 1737 4 O75097
23 107.5 6.5 1282 5 O18720
24 106 6.4 3010 12 Q9DTE7
25 105.5 6.4 911 11 Q9CRX6
26 105 6.4 984 5 Q9VLP7
27 104 6.3 773 5 Q9V514
28 104 6.3 773 5 O62335
29 103.5 6.3 398 11 Q9LYN8
30 103.5 6.3 1125 5 P92135
31 103.5 6.3 3010 12 Q9QIX2
32 103 6.3 552 15 Q83934
33 103 6.3 856 12 Q8QUT7
34 102.5 6.2 4123 4 O75851
35 102 6.2 420 5 P91776
36 102 6.2 552 15 Q83935
37 102 6.2 1026 5 Q8SMY0
38 102 6.2 2192 5 O01768
39 102 6.2 2327 13 Q9IBG7
40 101 6.1 552 15 Q83932
41 101 6.1 552 15 Q83933
42 101 6.1 1819 16 Q9ZLY0
43 101 6.1 2436 12 Q81756
44 101 6.1 3010 12 Q9QIX8
45 101 6.1 3010 12 Q9QIX7

```

ALIGNMENTS

RESULT 1

```

Q8VIE0 PRELIMINARY; PRT; 475 AA.
AC Q8VIE0;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Tesmin (Hypothetical 50.6 kDa protein).
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057423; BAB64935.1; -.
DR EMBL; BC024377; AAH24377.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF79F CRC64;

```

Query Match 76.0%; Score 1251; DB 11; Length 475;

Best Local Similarity 76.3%; Pred. No. 3.8e-122;

Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPQYQDQNNYLOSQDVPKPMWALVGRFLPASTK 60

```

|||||
181 MVICQLKGAQMCLICDNCARELKALHLLPQYDQSQSFQSELPKPMWTLVGRLLPVPAR 240

```

```
QY 61 LNLITQOOLGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCCNCCNNL 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 LNLITQVDNGALPSAVNGAFAFPGALQGPKITLSGYCDGFCDFSGDFCNCS ---CNL 296

QY 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVRPQHNGKNCRRSGCLKNKYCEYEAQI 180
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 RHELFRAIKACLRNPEAFQPKIGKQGLGNVRPQHNGKNCRRSGCLKNKYCEYEAQI 356

QY 181 MCSSTCKICGKNYEEPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 MCSSTCKICGKNYEEPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDRRPSS 416

QY 241 CISWEVVEATCACLAAQEEAEKHCSCKLAEQMILEFFGRLCSQILHTEPKSKGLKME 299
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 CISWEVVEATCACLAAQEEAEKHCSCSPSLAEQMLEFFGRLCSQILHTEPKSKGLKIE 475

RESULT 2
Q8VIE1
ID Q8VIE1 PRELIMINARY; PRT; 475 AA.
AC Q8VIE1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tesmin.
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057422; BAB64934.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Query Match 75.5%; Score 1244; DB 11; Length 475;
Best Local Similarity 75.9%; Pred. No. 2.1e-121;
Matches 227; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

QY 1 MVICOLKGGTOMLCIDNSTRRELKALHLVPOYQDNNYVQSDVPKPMALVGRFLPASTK 60
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 MVICOLKGGTOMLCIDNSTRRELKALHLVPOYQDNNYVQSDVPKPMALVGRFLPASTK 240

QY 61 LNLITQOOLGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCCNCCNNL 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 LNLITQVDNGALPSAVNGAFAFPGALQGPKITLSGYCDGFCDFSGDFCNCS ---CNL 296

QY 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVRPQHNGKNCRRSGCLKNKYCEYEAQI 180
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 RHELFRAIKACLRNPEAFQPKIGKQGLGNVRPQHNGKNCRRSGCLKNKYCEYEAQI 356

QY 181 MCSSTCKICGKNYEEPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 MCSSTCKICGKNYEEPERKTLMSMPNMTGGLEGSHYLPPTKFSGLPRFSDRRPSS 416

QY 241 CISWEVVEATCACLAAQEEAEKHCSCKLAEQMILEFFGRLCSQILHTEPKSKGLKME 299
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 CISWEVVEATCACLAAQEEAEKHCSPSLAEQMLEFFGRLCSQILHTEPKSKGLKIE 475

RESULT 3
Q9V6Q8
ID Q9V6Q8 PRELIMINARY; PRT; 950 AA.
AC Q9V6Q8;
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Folsler C., Gabrielian A.E., Garg N.S., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Mays A.D., Dew I., Dietz S.M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stadling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003818; AAF58365.1; -.
DR FlyBase; FBgn0033846; CG6061.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 950 AA; 100021 MW; 627C3EA6A44A0A30 CRC64;

Query Match 28.5%; Score 470; DB 5; Length 950;
Best Local Similarity 34.9%; Pred. No. 4e-40;
Matches 104; Conservative 47; Mismatches 81; Indels 66; Gaps 7;

QY 30 PQYQDNNYVQSDVPKPMALVGRFLPASTKLNLIITQOL-----EGALPSV---- 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 PQYQSKHQQLQ-----AQAKRIQQOQLPTEQSTPIKVEKPLTLPVG 703
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 76 -VNGSAFPGSTLPGPKITLAG-----AQAKRIQQOQLPTEQSTPIKVEKPLTLPVG 109
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 704 KANVPKPLFEVLKPPATAAAGAVDPLGGMTSRKHCNCSKSOCLKLYCDFANGFECQ 763
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 110 NCNCCNCCNLLHDDIERFAIKACLRNPEAFQPKIGKQGLGNVRPQHNGKNCRRSGCL 169
|:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 764 DCTCKDCFNLLDYEVERERAIRSCLDRNPFAKPKITAPNSGDMR-LHNKGCNCRSGCL 822
```



```
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;

Query Match 22.8%; Score 376; DB 10; Length 571;
Best Local Similarity 29.4%; Pred. No. 1.4e-30;
Matches 95; Conservative 48; Mismatches 90; Indels 90; Gaps 10;

QY 41 SDVPRMTALVGRFLPASTKLNLT-----QOLEGAL-----PSVNGSAPPSGST 86
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 51 ASIPSPITVTRPITTSQAPTVATPIPPPSQSGIILHVPTRPSPSPMRPAGET 110
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 87 LPGPPK-----ITLAGYDCFCASGDFCNCCNCCNHHDIERFKAKKACLAGR 136
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 111 RDGTPKKKQCCKHSRCKLYKCECFASGTGCDGNCVCFNNVFNPAQVETLER 170,
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 137 NPEARQPKI-----GKQGLNV--KPOHNGKCNRRSGCLKNKYCECEYAOIMCSCI 185
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 171 NPNAFRPKIAASPHGGRDNREVGDVVMLARHNKGCHCKGKLYKCECFQANILCSEN 230
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 186 CKICGKNYKESPERKTLMSPN-----YMQ-----TGGLEGSHYL---PPTKFSG--- 228
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 231 CKCLCKNFESGEVRSQFLFHEGSHSNLAYLOHANAATGATGSGGFASAPPPKRRKGQEI 290
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 229 -----LPRFSDRRPSS-----CISWEVWEAT 250
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 291 PFNQGTSDSSTRHQANNGRTTSOTGSRAGNASLGPVKVYKSLLANIIPKMDVKAL 350
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 251 CACLLAQGEAEKEHCSCLAEO 273
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 351 CSVLVAVAGEAAKLTLEKRLANQ 373
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 7
Q9D571 PRELIMINARY; PRT; 277 AA.
AC Q9D571;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4930509C02, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J. H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK015732; BAB29949.1; -.
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;
```

```
Query Match 22.4%; Score 369; DB 11; Length 277;
Best Local Similarity 73.2%; Pred. No. 3.1e-30;
Matches 71; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 1 MVICQLKGGTOMLCIDNSRTRELKALHLVPOYQDQNNYLOSDVQPKPMTALVGRFLPASTK 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 MVICQLKGGTOMLCIDNCGARELKALHLPOYDQSSPQSELPKPMPTTLVGRLLPVPK 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 61 .LNLITQOLEGALPSVWNGSAPPSGTLPPKPIITLAG 97
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 241 LNLITQVDNGALPSAVNGAAPPSPALQGPPIITLSG 277
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 8
Q9SZD1 PRELIMINARY; PRT; 603 AA.
AC Q9SZD1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcription factor.
GN F19B15.30 OR AT4G29000.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078470; CAB43914.1; -.
DR EMBL: AL161574; CAB79658.1; -.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AFE818DE5 CRC64;

Query Match 21.7%; Score 357; DB 10; Length 603;
Best Local Similarity 28.2%; Pred. No. 1.5e-28;
Matches 96; Conservative 43; Mismatches 83; Indels 118; Gaps 12;

QY 39 LQSDVP---RP-MTALVGRF-----LPATKLNLTITQOLEGALPSVWNGSAPPSGTL 88
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 58 ISSVSPSTIRPGMTIAIGQTVQVTRTPMATTPSPSQS-----IVNA---FIRPIP 109
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 89 GPPKI-----TLAGYDCFCASGDFCNCCNCCNCCNLLH 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 110 ESPRARGPRNVEGRDGTPOKKKOCNCKHSRCKLYKCECFASGTGCDGNCVCFNNVDN 169
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 123 DIERFKAICACLGNNPEAFQPKICKGOLGN-----VKPOHNGKCNRRSGCLKN 171
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 170 EPARREAVEATLENPAFRPKIASSPHGGDRDKREDIGEVLLCKHNKGCHCKSGCLKK 229
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 172 YCEYEQAQIMCSSICKICGKNYKESPERKTLMSPN-----YMQ-----TGGLEGSH 219
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 230 YCECFQANILCSENCKLDCCKNFEGSEERQALFHEGSHNIMAYLQQAANAATGAVGSSG 289
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 220 YLP-----PTKFSGLPRFSDRR--PSS----- 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```


DE Putative DNA binding protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY046019; AAK76693.1; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 229.5; DB 10; Length 526;
Best Local Similarity 24.2%; Pred. No. 2.7e-15;
Matches 55; Conservative 31; Mismatches 72; Indels 69; Gaps 5;

Qy 98 YDCDFASGDFC-NNCNCNNCCNHHDIERFKATKACLRNPEAFQPKI-----145
Db 245 YCECFAAGVYCEPCSDICDFNKPIHEETVLATRKQIESRNPLAPAPKVI RNADSIMEAS 304
Qy 146 GKGQLGNVKKPOHNGCNCRRSGCLKNKYCEYEAQIMCSICKICGCKNY-----194
Db 305 DDASKTPASARHKRCNCCKSNCKKCYCYGGVGCSSMNCRCGCTNVFGKDGSLVI 364
Qy 195 -----EESPERKTLMSMPNYMOTGGL-----GSHYL 221
Db 365 MESKLEENQETYEKRIAKIOHNVEVSKEVEQNPPSSDQSTPLPPYRHLVWHQPFSLKNRL 424
Qy 222 PPTKPF-----SGLPRFSDHRRPSSCISWEVVEATCACL 255
Db 425 PPTQFFLGTGSSFRKPNSDLAQSQNEKKPLETVTDKTEIMPEILL 471

RESULT 15
Q9LE32
ID Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TS01 (Putative DNA binding protein).
GN TS01.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. LANDSBERG;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TS01 is a novel protein that modulates cytokinesis and cell expansion
in Arabidopsis.";
RL Development 127:2219-2226(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
Arabidopsis gene with cysteine-rich repeats.";

RL Development 127:2207-2217(2000).
DR EMBL; AF204059; AAF69124.1; -;
DR EMBL; AF206324; AAF27433.1; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Query Match 13.9%; Score 229.5; DB 10; Length 695;
Best Local Similarity 24.2%; Pred. No. 3.8e-15;
Matches 55; Conservative 31; Mismatches 72; Indels 69; Gaps 5;

Qy 98 YDCDFASGDFC-NNCNCNNCCNHHDIERFKATKACLRNPEAFQPKI-----145
Db 414 YCECFAAGVYCEPCSDICDFNKPIHEETVLATRKQIESRNPLAPAPKVI RNADSIMEAS 473
Qy 146 GKGQLGNVKKPOHNGCNCRRSGCLKNKYCEYEAQIMCSICKICGCKNY-----194
Db 474 DDASKTPASARHKRCNCCKSNCKKCYCYGGVGCSSMNCRCGCTNVFGKDGSLVI 533
Qy 195 -----EESPERKTLMSMPNYMOTGGL-----GSHYL 221
Db 534 MESKLEENQETYEKRIAKIOHNVEVSKEVEQNPPSSDQSTPLPPYRHLVWHQPFSLKNRL 593
Qy 222 PPTKPF-----SGLPRFSDHRRPSSCISWEVVEATCACL 255
Db 594 PPTQFFLGTGSSFRKPNSDLAQSQNEKKPLETVTDKTEIMPEILL 640

Search completed: April 21, 2003, 11:49:59
Job time : 30.1886 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:19:52 ; Search time 95.879 Seconds
(without alignments)
6228.992 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 4008
Sequence: 1 tatcctgtggtggcccg.....aaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cygn2_1/USPTO_spool/US09743237/runat_21042003_111943_381/app_query.fasta_1.4494
-DB=A_Geneseq_101002 -QEMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALL=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -SCGN_1_1_259 -runat_21042003_111943_381 -NCPUP=6 -ICPU=3
-NO_XLPYX -NO_MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	39.9	295	21	AAV68463
2	1251	31.2	299	21	Mouse testis speci
3	503	12.5	147	22	Human testis speci
4	503	12.5	147	22	Human testicular a
5	440.5	11.0	950	22	Human reproductive
6	405.5	10.1	438	20	Drosophila melanog
7	362	9.0	280	22	Caenorhabditis ele
8	346	8.6	251	22	Novel human diagno
9	248	6.2	53	22	Human polypeptide
10	186	4.6	243	22	Human colon cancer
11	140.5	3.5	4561	22	Drosophila melanog
12	140.5	3.5	9222	22	Novel human diagno
13	136.5	3.4	966	23	Novel human diagno
14	131.5	3.2	1061	20	Human TSP1 domain
15	127	3.1	961	23	Human N-methyl-D-a
16	126.5	3.2	1454	22	Human TSP1 domain
17	125	3.1	961	23	Human polypeptide
18	124.5	3.1	1081	23	Human TSP1 domain
19	124.5	3.1	1081	23	Human N-methyl-D-a
20	124	3.1	3680	22	Human NMDA recepto
21	122.5	3.1	1017	22	Drosophila melanog
22	122.5	3.1	1615	22	TutD protein #4.
23	120.5	3.0	966	23	Protein #3 encoded
24	120.5	3.0	1212	20	Human TSP1 domain
25	119.5	3.0	386	22	Human N-methyl-D-a
26	119.5	3.0	1185	20	Human purified sec
27	119.5	3.0	2594	16	Human atrophin I p
28	119.5	3.0	3541	23	IgG-Fc binding pro
29	119.5	3.0	5405	16	Human melanoma spe
30	119.5	3.0	7337	22	IgG-Fc binding pro
31	118.5	2.9	1413	23	Novel human diagno
32	118.5	3.0	1422	22	Human PKIN-24 prot
33	118.5	3.0	3118	22	Novel human diagno
34	117.5	2.9	732	18	Human SRCAP. Homo
35	117.5	2.9	2972	22	Human RECK cancer-
36	117	2.9	566	22	Human SRCAP. Homo
37	117	2.9	1028	21	Human polypeptide
38	115.5	2.9	1033	22	A receptor-like pr
39	114	2.8	1082	23	Novel human diagno
40	114	2.8	2157	21	Human polypeptide
41	113.5	2.8	1232	21	A human hyaluronan
42	113.5	2.8	1232	21	Human N-methyl-D-a
43	113.5	2.8	2971	21	Human NMDA recepto
44	113	2.8	349	21	Human ORFX ORF995
45	113	2.8	746	22	Mouse taste cell s
					Drosophila melanog

ALIGNMENTS

RESULT 1
AAV68463

ID AAV68463 standard; Protein; 295 AA.

AC AAV68463;

DT 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.

OS Mus musculus.

XX WO200004147-A1.

XX 27-JAN-2000.

XX 16-JUL-1999; 99WO-JP03859.
XX 17-JUL-1998; 98JP-0219856.
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX WPI: 2000-147785/13.
XX N-PSDB; AA288155, AA288157.
XX New male germ cell regulatory factor tesmin expressed in spermatocytes
XX useful for investigation of germ cell differentiation and sterility -
XX Claim 1; Page 47-49; 63pp; Japanese.
XX The present sequence represents a male germ cell regulatory factor
XX expressed specifically in spermatocytes, designated tesmin. Tesmin
XX can be used in the investigation of the mechanisms of germ cell
XX differentiation and sterility.
XX Query Match: 39.90%
XX Indels: 0
XX Gaps: 0
XX Sequence 295 AA;
Alignment Scores:
Pred. No.: 2,79e-142 Length: 295
Score: 1599.00 Matches: 295
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.90%
DB: 21
US-09-743-237-1 (1-2241) x AAY68463 (1-295)
QY 651 ATGGTATTTGTCTAGCTGAAGAGGGCCCGCATGCTCTGCATAGACACTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 711 AGGAGCTCAAGCGCTCCATCTGCTCCCTCAGTACCATCAGCAGACAGCTTTCCCTCAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGln 40
QY 771 TCAGAGCTCCCTAAGCCCAATGACAACTTTAGTGGGAGACTTCTGCCAGTACCAGCGAAG 830
DB 41 SerGluLeuProLysProMetThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 831 TTAATCTCATCACAGGTTGATATGGAGCTCTCCCATCAGCTGCTCAATGGGGTCC 890
DB 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 891 TTTCCCTCTGGACTGCTCTGCAAGGCGCCACCCAAATAACTCTGCTGGGTACTGTGAC 950
DB 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTGCGCCCATGAGCTC 1010
DB 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnLeuArgHisGluLeu 120
QY 1011 GAGCGTTCACAGCCATAAAGCGGTGTCTTGATAGAAATCCTGAAGCTTTCCACCAAAA 1070
DB 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
QY 1071 ATGGGAAAGCGCGCTGGGAGCTGTAACTTCGACACAGCAAGGCTGCAACTGTAAG 1130
DB 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
QY 1131 CGCTCAGGCTGCCTGAAGAACTACTGTGAGTGTATGAGGCCAAATCATGTGTCTTCC 1190
DB 161 ArgSerGlyCysLeuLysAsnTyrCysGlyCysTyrGluAlaLysIleMetCysSerSer 180
QY 1191 ATTTGCAATGCAATGCTTCCAAATACTATGAGAAAGTCCAGAACCAAAATGCTGATG 1250
|||||

DB 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
QY 1251 AGCACCCACATACATGAGCCTGGGACTTTGAGAGCAGCAGCATTTTGTCCCGAGCC 1310
|||||
DB 201 SerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuSerProAla 220
QY 1311 AGTTCTCAGGACCTCCAAAACCTAGAGAAAATAGGAGGCGCTTCTCCTGTATCTCCTGG 1370
|||||
DB 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTrp 240
QY 1371 GAAGTAGTGGAGGCCACATGCTGCTGCTGCTGCCAGGCTGAGGAGGAGGAGGAGGAG 1430
|||||
DB 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaGluGlnGlu 260
QY 1431 CACTGTTCCTCCCAAGCTTGGCTGAGCAGCATCTCTGAGGAGGAGTTTGAAGGTCCTGTCTG 1490
|||||
DB 261 HisCysSerProSerLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSer 280
QY 1491 CAGATTCTCCATCAGTTCAGTTCAGTTCAGGCGCTGAAATTTGAG 1535
|||||
DB 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
RESULT 2
AAY68464
ID AAY68464 standard; Protein; 299 AA.
XX
AC AAY68464;
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-JP03859.
XX
PR 17-JUL-1998; 98JP-0219856.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX
DR WPI: 2000-147785/13.
DR N-PSDB; AA288156.
XX
PT New male germ cell regulatory factor tesmin expressed in spermatocytes
XX useful for investigation of germ cell differentiation and sterility -
XX
PS Claim 1; Page 50-52; 63pp; Japanese.
XX
CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin. Tesmin
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 299 AA;
Alignment Scores:
Pred. No.: 2,47e-109 Length: 299
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 31.21% Indels: 4
DB: 21 Gaps: 1

KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX Homo sapiens.
OS WO200155320-A2.
XX 02-AUG-2001.
PN 17-JAN-2001; 2001WO-US01339.
PD 31-JAN-2000; 2000US-0179065.
PF 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

Db 656 ysGlnMetGlyPro-GlnGlnSerLys-----HisGlnGlnLeuGln 670
 Qy 665 GCTGAAGAGGCGCCAGAGTCTGCATAGACAACAGTGGCGGAGGAGCTCAAGC 724
 Db 671 AlaGlnAlaLysGln-----ArgIleArgGlnGlnGlnLeuProThrGluGlnSer 687
 Qy 725 GCTCCATCTGCTTCTCAGTACGATACGACGAGCAGTTCCTCAGTCAGAGCTCCCTAA 784
 Db 688 ThrProIleLysValGluProLysLeuProThrLeuProGlyValLysAla-AsnVa 707
 Qy 785 GCCAATGACAACTTGTAGTGGAGACTTCTGCCAGTACACGGAAGTAAATCTCATCAC 844
 Db 707 IProAlaLysProLeuPheGluValLysPro---ProAla----- 720
 Qy 845 ACAGGTTGATATGAGCTCTCCCATCAGCTCAATGGGCTGCTTCCCTCTCGGACC 904
 Db 721 -----ThrAlaAlaAlaGlyAlaValAspProLeuGlyGln 733
 Qy 905 TGCTCTGCAAGGCGCCACCAATAACTCTCTCTGGG-----TACTG 946
 Db 733 yMetThrSerArgArgLysHisCysAsnCysSerLysSerGlnCysLeuLysLeuTyrCy 753
 Qy 947 TGACTGCTTCTCCACGGGGACTTCTGCAACAGCTGCAGCTGC-----AACAA 994
 Db 753 sAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLysAspCysPheAsnAs 773
 Qy 995 CTGCGCCATGAGCTCGAGCGCTTCAAGCCATAAGCGGTGCTGTATAGAAATCCCTGA 1054
 Db 773 nLeuAspTyrGluValGluArgGluArgAlaIleArgSerCysLeuAspArgAsnProSe 793
 Qy 1055 AGCTTTTCCAAACAAATGGGAAAGCGCTGCTGGAGCTGTAACTTCGACAGCA 1114
 Db 793 rAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetArgLeu---HisAsn 812
 Qy 1115 AGGTGCAACTGTAAAGCGCTCAGCGCTGCTGAAGAACTACTGTAGTGTATGAGGCCAA 1174
 Db 812 sGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLys 832
 Qy 1175 AATCATGTTCTTCCATTTCGAATGATTGCTTGCAGAACTATGAAGAAGTCCAGA 1234
 Db 832 sIleProCysSerSerIleCysLysCysValCysArgAsnMetGluAspArgProAs 852
 Qy 1235 AGAAAAATGTGTATGATGACACACCCCTACATGAGCGCTGGGGACTTTGAGAGCGCCA 1294
 Db 852 p-----ValAspMetAspSerLeuAs 859
 Qy 1295 TTATTTTCCCGCCAGCAAGTCTCAGGAGCTCCAAACTGAGAAAAATAGGAG- 1349
 Db 859 pGlyLeuMetGlyValGluGlyGlnLysLysAspLys---AlaLysAsnLysGlnLeuAs 878
 Qy 1350 -----GCCTTCTCTGTATCTCTGGAAGTACTGAGGCGCCACATGTGCTGCT 1399
 Db 878 nGluAsnArgAlaAsnIleTyrPheThrAspAspValIleGluAlaThrIleMetCysMe 898
 Qy 1400 GCTGCGCCAGGGTGGAGGAGCAGCAGCAGTGTTCCTCCCAAGCTTGGCTGAGCAG 1459
 Db 898 tIleSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGln 918
 Qy 1450 GATCTGTGAGAGGTTTGAAGGTGCTGTGCGAGATTCTC 1499
 Db 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931
 RESULT 6
 ID: AAW83392 standard; Protein: 438 AA.
 XX
 AC AAW83392;
 XX
 DT 29-MAR-1999 (first entry)
 XX
 DE Caenorhabditis elegans synMuv protein LIN-54.
 XX
 KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;

KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 KW cell proliferation; gene therapy.
 XX Caenorhabditis elegans.
 OS WO9854299-Al.
 PN 03-DEC-1998.
 PD 28-MAY-1998; 98WO-US11043.
 PF 28-MAY-1997; 97US-0047996.
 PR (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA Ceol C, Horvitz HR, Lu X;
 PI WPI; 1999-045362/04.
 DR N-PSDB; AAV72865.
 XX Novel LIN-37, -35, -55, -52, and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas
 PS Claim 7; Fig 13; 70pp; English.
 XX

This is the amino acid sequence of LIN-54, a novel protein of
 Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 synthetic multivulvar (synMuv) gene involved in cell fate and cell
 proliferation, and is part of a pathway that may be used as a
 genetic and biochemical model system for tumour suppression and
 cancer in mammals. SynMuv pathway genes and proteins may be used
 to identify genes which are part of the mammalian pathway and
 to identify genes, proteins and therapeutic compounds which
 modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 polypeptide, and (5) an antibody which binds to a SynMuv family
 protein. The SynMuv nucleic acids and polypeptides can be used to
 diagnose and treat, especially by gene therapy, conditions
 involving altered levels of cell proliferation, e.g. SynMuv-
 associated carcinomas.

SQ Sequence 438 AA;

Alignment Scores:
 Pred. No.: 3.25e-29 Length: 438
 Score: 405.50 Matches: 132
 Percent Similarity: 44.37% Conservative: 57
 Best Local Similarity: 30.99% Mismatches: 158
 Query Match: 10.12% Indels: 79
 DB: 20 Gaps: 18

US-09-743-237-1 (1-2241) x AAW83392 (1-438)

Qy 419 CAGGGGAAATCATTTTGGTGAAGTCCGATGAACACTGTCAGGTCCCAAGCAGCAA 478
 Db 3 GlnGlyGluIleVal-TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22
 Qy 479 GCAACCAAGAAAAGTGTGAATCAAGACGAGTGTAGTGTGCCAGGCGGAGCC 538
 Db 22 pAsnTyrGluGluGlyAlaGluPheIleGluValAsnGlyGlnLeuValProHisAsnPr 42
 Qy 539 TGAGACGACGCTTTCAGGCGCCCTCTGCTCAGGAATCTGTTGCAAGTTCCTCATCATC 598
 Db 42 O-AsnLeuGlnAlaGlnGln-----AsnArgProGlyThrSerMetIleG 58
 Qy 599 CCAGGAGCAGAGGAGCGCTCCAGCTGCCCTCGGAAGAAGAACTCCAGCCCCATGGTAT 658
 Db 58 InGlnHisAsnArgSerMetGluValAsnGlnGlyLeuValLysAspGluProIle-Asp 77

QY	659	TTGTCAGCTGAAGG-----AGGCCCCAGATGCTCTGCATAGACAACACTGTGGCGC	709
Db	78	ThrSerHisargValTyrValProProArgProValGlnArgLysLeuTriplys	97
QY	710	----GAGGAGCTCAAAAGCGTCCATCTGCTTCCCTCAGTAC-----GA	748
Db	98	LeupheGlnProGlyProserThrProGlySerSer-GlnTyrThrValArgAsnLeuSe	117
QY	749	TGACCAGAGCAGGTTTCCCTCAG-----TCAGAGCTCCCTTAAG-----	785
Db	117	rAsnLeuSerGlySerProSerMetTyrAspArgGlnProAlaSerLeuProArgTrVa	137
QY	786	----CCAAATCACAACTTTAGTGGAAGACTTCTCCAGTACCAGGCAAGTTAAATCTCAT	841
Db	137	lGlnProMetGlyLeuGluMetGlyAsnSerGluGlnArgLysValTyrIleAspMetLy	157
QY	842	CACACAGGTTGATAAATGGAGCTTCCCATCAGCTGTCAATGGGCGTGCTTCCCTCGG	901
Db	157	sAspHisValSerHisIleargLeuLysThrIlysLysValPheAla-----	173
QY	902	ACCTGCTGTCGAAGGCCACCCAAAATAACT-----CTGCTCGGTACTGTGA	949
Db	174	-ProGlyGlnArgLysProCysAsnCysThrIlysSerGlnCysLeuLysLeuTyrCysAs	193
QY	950	CTGCTTCTCCAGCGGAGCTTTCGCAACAGCTGCAGCTGC-----AACACCT	997
Db	193	pCysPheAlaAsnGlyGlupHeCysArgaspCysAsnCysLysaspCysHisAsnAnil	213
QY	998	GCGCATAGCTCAGCGGCTTCAAAGCCATAAAGCGTCTTGATAGAAATCTCTGAAGC	1057
Db	213	eGluTyrAspSerGlnArgSerLysAlaIleargGlnSerLeuGluArgAsnProAsnAl	233
QY	1058	TTTCCAACAAAAATGGGAAGCGCGTCTGGAGCTGCTAAACTTGA-----CACAG	1111
Db	233	aPheLysProLysIleclylealaArgglyIlethrAspIleGluArgLeuHisGl	253
QY	1112	CAAGGGTGCAACTGTAAAGCGCTCAGGTCGCTGAAGAAGCTACTGTGAGTCTATGAGGC	1171
Db	253	nLysGlyCyHisCysLysSerGlyCysLeuLysasnTyrCysGluCyStyrgLual	273
QY	1172	CAAAATCATGTGTTTCCATTGCAAAATGCATTGCTGCTCAAAAC-----	1217
Db	273	aLysValProCysThrAspArgCysLyscysLysGlyCysGlnAsnThrGluThrTyAr	293
QY	1218	-----TATGAAGAAGT-----CCAGAAGCAAAAATGCTGATGAGC--	1253
Db	293	gMetThrArgTyrLysAsnSerGlyglyAlaValSerAsnThrAsnAlaLeuMetSerPh	313
QY	1254	-----ACACCCACATACATGGAGCGCTGGGAGCTTTTGAGACGAC	1291
Db	313	eThrAsnAlaSerSerThrAlaThrProasp---SerGlyproGlySerValValThrAs	332
QY	1292	CCATTAT-----TTGTCGCCAGCCAAGTTCTCAGGACCTCCAAACCTGAGAAAAA	1342
Db	332	pGluHisGlyAspAspTyrGluaspMetLeuLeuSerHisLysProLysValGluMetAs	352
QY	1343	TAGCAGGCGCTTCTCTGTATCTCTCG-----GAAGTAGTGGAGGCCAC	1387
Db	352	pProArgProphe-----ProTprTyrTyrMethThrAspGluValValGluAlaAl	369
QY	1388	ATGTCCTGCTGCTGCGCCACAGGGTGAGGAACA-----GAGCAGGA	1429
Db	369	aThrMetCysMetValAlGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGl	389
QY	1430	GCATGTGTCCCCAAGCTTGGCTCAGCAGATGATCTCGGAGGAGTTTGAAGGTGCCTGTC	1489
Db	389	uaspgLuLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGl	409
QY	1490	GCAGATTCTC	1499
Db	409	uGlnMetile	412

RESULT 7

RESULTS /
ABG17958
ID ABG17958 standard; Protein; 280 AA.
XX
AC ABG17958;

DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17949.

Human; chromosome mapping; gene mapping; gene therapy; forensic; KW
food supplement; medical imaging; diagnostic; genetic disorder; KW
food supplement; medical imaging; diagnostic; genetic disorder; KW

OS Homo sapiens.

XX WQ200175067-A2
PN

XX
PD
11-OCT-2001XX
PF
30-MAR-2001. 2001WO-IIS0863131-MAR-2000. 2000US-0540217
XX
PP

PR 31-MAR-2000; 2000US-0340Z17-
PR 23-AUG-2000; 2000US-0649167-

XX
PA (HYSE-) HYSEO TNC

XX	PT	Dymanac	PT	Liu C.	Yang Y.T.

XX
DR
WPT: 2001-639362/73

DR WPI; Z001-63936Z/
DR N-PSDB: AAS82145

xx New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20: SEO ID No 48317: 103pp: English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABC00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

XX	Sequence	280 AA:
SO		

Alignment Scores:

Alignment Scores:	3.45e-25	Length:	280
Pred. No.:	362.00	Matches:	66
Score:	71.05%	Conservative:	15
Percent Similarity:	57.89%	Mismatches:	23
Best Local Similarity:	9.03%	Indels:	10
Query Match:	22	Gaps:	2
DB:			

US-09-743-237-1 (1-2241) x ABG17958 (1-280)

942 TACTGTGACTGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCAGCTGC----- 989


```
|||||
Db 114 TyrCysAspCysPheAlaAsnGlyGluPheCysAsnAsnCysAsnCysThrAsnCysTyr 133
QY 990 AACAACTGCCCGATGAGCTGAGCGCTTCAAAGCCATAAAGGGGTGCTGTGATAGAAAT 1049
Db 134 AsnAsnLeuGluHlaGluAsnGluArgGlnLysAlaTleLysAlaCysLeuAspArgAsn 153
QY 1050 CCTGAAGCTTCCCAACCAAAATGGGAAAGCGCTCTGGGAGCTGCTAAACTTCGACAC 1109
Db 154 ProGluAlaPheLysProLysIleGlyLysGlyLysGluGlyGluSerAspArgArgHis 173
QY 1110 AGCAAGGTCGCAACTGTAAAGCTGCTGAGCTGCTGAGAACTACTGTGAGTGTATGAG 1169
Db 174 SerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGlu 193
QY 1170 GCCAAATCATGCTGCTTCCATTGCAATGCAATGCAATGCTTCCAAAACTATGAGAAGT 1229
Db 194 AsnIlePheIleTrpTyrCysCysArg-----TyrArgGlnIle 207
QY 1230 CCAGAACAAAATGCTGATGAGCAGACACCCACATACATGGAG 1271
Db 208 LeuGluLysAlaIleGlnLeuSerGlyAlaGluGlnLeuGlu 221
RESULT 8
AAM93348
ID AAM93348 standard; Protein: 251 AA.
XX AC
XX AAM93348;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide, SEQ ID NO: 2895.
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX OS Homo sapiens.
XX PN EP1130094-A2.
XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-0114089.
XX PR 08-JUL-1999; 99JP-0194486.
XX PR 11-JAN-2000; 2000JP-0118774.
XX PR 02-MAY-2000; 2000JP-0183765.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX DR WPI; 2001-524255/58.
XX DR N-PSDB; AAK94268.
XX PT 830 Primers useful for synthesizing full length cDNA clones and their
XX PT use in genetic manipulation -
XX PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX CC The invention relates to primers for synthesizing full length cDNA
XX CC clones. 830 cDNA molecules encoding a human protein have been
XX CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX CC molecules have been determined. Primers for synthesizing the full length
XX CC cDNA are useful for clarifying the function of the protein encoded by
XX CC the cDNA. The full length clones were obtained by construction of full
XX CC length enriched cDNA libraries that were synthesised by the oligo-capping
XX CC method. The primers enable the production of the full length cDNA easily
XX CC without any special methods. The present sequence is a polypeptide
XX CC encoded by a full length human cDNA of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in CD-ROM format directly from EPO.
```

```
SQ Sequence 251 AA;
Alignment Scores:
Pred. No.: 1.07e-23 Length: 251
Score: 346.00 Matches: 66
Percent Similarity: 80.65% Conservative: 9
Best Local Similarity: 70.97% Mismatches: 18
Query Match: 8.63% Indels: 0
DB: 22 Gaps: 0
US-09-743-237-1 (1-2241) x AAM93348 (1-251)
QY 495 GTTGAATCAAGAAGCAGGTGGTAGTGTGCCAGCGCGACCCCTGAAGACGACGCTTTC 554
Db 158 ValGluIleLysGluAlaGlyClyThrThrSerAsnAsnProGluGluAlaThrLeu 177
QY 555 CAGCGCCCTCTGCGTCAAGAACTCTGTTGCAAGTTCATCCATCCAGGAGGAGGAG 614
Db 178 GlnAsnLeuLeuAlaGlnLysCysCysLysPheProSerSerGlnGluLeuGluAsp 197
QY 615 GCCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTGTGTCAGCTGAAGGA 674
Db 198 AlaSerCysCysSerLeuLysLysAspSerAsnPrometValIleCysGlnLeuLysGly 217
QY 675 GCGCCAGATGCTCTGCATAGACAACTGTGGCGCGAGGAGCTCAAGAGCCTCCATCTG 734
Db 218 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgGluLeuLysAlaLeuHisLeu 237
QY 735 CTTCCTCAGTACGATGACCGAGCAGTTCCTCCCTCAGTCA 773
Db 238 ValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSer 250
RESULT 9
AAG76158
ID AAG76158 standard; Protein: 53 AA.
XX AC
XX AAG76158;
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:6922.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma; chromosome 11.
XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX DR WPI; 2001-235357/24.
XX DR N-PSDB; AAH35563.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PS Claim 11; Page 8365; 9803pp; English.
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
```

CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 53 AA;

Alignment Scores:

Pred. No.:	9.71e-15	Length:	53
Score:	248.00	Matches:	48
Percent Similarity:	94.34%	Conservative:	2
Best Local Similarity:	90.57%	Mismatches:	3
Query Match:	6.19%	Indels:	0
DB:	22	Gaps:	0

US-09-743-237-1 (1-2241) x AAG76158 (1-53)

QY 1377 GTGGAGGCCACATGTCCTGCTGCTGCGCCAGGCTGAGGAAGCAGCAGGAGCACTCT 1436
Db 1 ValGluAlaThrCysAlaCysLeuAlaGlnGlyGluAlaGlnLysGluHisCys 20
QY 1437 TCCCAAGCTTGCTGAGCAGATGATCCCTGGAGGAGTTTGGAGGTCCTGTGCAGATT 1496
Db 21 SerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSerGlnIle 40
QY 1497 CTCACATGATTCAGTCCAGGCCAGGCTGGAATTTGAG 1535
Db 41 LeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 53

RESULT 10

AB68888
ID AB68888 standard; Protein; 243 AA.

XX
AC AB68888;

XX
DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.

XX
PR 11-JUL-2000; 2000US-0614150.

XX
PA (PEXE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;

XX
DR WPI; 2001-656860/75.

XX
DR N-PSDB; ABL12991.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX

PS Disclosure; SEQ ID NO 33456; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB72072).

CC

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 243 AA;

Alignment Scores:

Pred. No.:	1.48e-08	Length:	243
Score:	186.00	Matches:	59
Percent Similarity:	36.64%	Conservative:	26
Best Local Similarity:	25.43%	Mismatches:	61
Query Match:	4.64%	Indels:	86
DB:	22	Gaps:	8

US-09-743-237-1 (1-2241) x AB68888 (1-243)

QY 1044 AGAATCTCTGACGTTTCCCAACCAAAATGGGAAGGCCGCTGGAGCTGCTAAACTT 1103
Db 7 ArgSerValAspLysAlaAspGlyLysGlyGln---GlyAlaGlyVal 25

QY 1104 CGACACAGCAAAAGGGTGCAACTGTAGCGCTCAGCGCTGCTGAAGAACTACTGTAGTGC 1163
Db 26 -----LysGlyCysCysCysLysArgSerGlnCysIleLysAsnTyrCysAspCys 42

QY 1164 TATGAGGCCAAATCATGTGTTCTTCATTTGCAATTCGATTCGCAAAACATATGAA 1223
Db 43 TyrGlnSerMetAlaIleCysThrLysPheCysArgCysValGlyCysArgAsnThrGlu 62

QY 1224 -----GAAAGTCCAGACCA--- 1238
Db 63 ValArgGluLeuValAspProAsnSerValAlaLysAsnSerSerAlaValLysArgGln 82

QY 1238 ----- 1238

Db 83 LysAlaAlaAlaMetSerAlaLysAlaAlaAlaAlaAlaLysAlaGlyIleAspVal 102

QY 1239 -----AAAATGCTGATG 1250
Db 103 GlnGlyLysAlaLeuGlnValAlaAlaSerThrLeuAlaLeuProGlyLysAlaLeuMet 122

QY 1251 AGCACACCCCACTACATG-----GAGCCCTGGGAGCTTTCAGAGCAGCCATTAT 1298
Db 123 ThrProProLysTyrThrLeuValAlaGlyLysPro---ProMetAlaSerSerHis--- 140

QY 1299 TTGTCCCCAGCCCAAGTTCTCTCAGGACCT----- 1325
Db 141 IleAsnProIleProIleSerArgProIleAlaThrAlaAlaThrProAlaArgAlaVal 160

QY 1326 -----CCAAACTGAGAAAAAT 1343
Db 161 LysGlnProAlaGluProProMetProValAsnLeuIleIleProValArgHisAspAsp 180

QY 1344 AGGCAGGCTTCTCCTGTATCTCCTGGGAAGTAGTGGAGGCCACATGTGCTCGCTGCTG 1403
Db 181 ArgArgAspArgAsnLeuPheValGlnProValAsnAlaAlaLeuLeuGluCysMetLeu 200

QY 1404 GCCCAGGTGAGGAAGCAGCAGCAGGAGCAGCTGTTCCCAAGCTTGGCTGAGCAGATGATC 1463
Db 201 IleGlnAlaThrGluAlaGluGlnLeuGlyLeuAsnGluLeuGlnValCysGlnLeuVal 220

QY 1464 CTGGAGGAGTTTGGAGGTGCTGTCGACGATTC 1499
|||||

Db 221 LeuGluGluPheMetArgGlyTyrLysAsnIleLeu 232

RESULT 11

ABG30203
ID ABG30203 standard; Protein; 4561 AA.

XX AC ABG30203;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #30194.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX PT WPI; 2001-639362/73.

XX DR N-PSDB; AAS94390.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX PS Claim 20; SEQ ID No 60562; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.XX CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 4561 AA;

Alignment Scores:

Pred. No.:	0.00121	Length:	4561
Score:	140.50	Matches:	140
Percent Similarity:	30.35%	Conservative:	57
Best Local Similarity:	21.57%	Mismatches:	194
Query Match:	3.51%	Indels:	258
DB:	22	Gaps:	40

US-09-743-237-1 (1-2241) x ABG30203 (1-4561)

QY	516	GGTAGTGTGCCAGGCGGCAGCCTTGAAGAGCGCAGCTTTCCAGCGCCCTCTGGCTCAGGAA	575
Db	1772	GlyThrAlaProProAlaAlaPheGlyGlyAlaGluCysGlnGlyProThrMetGluAla	1791
QY	576	TCCTGTTGCAAGTTC-----CCATCATCCAGGAGCAGAGGAGGCGCTCCAGCTGC	626
Db	1792	GlupheCysSerLeuArgProCysProAlaSerTrpGlyCysCysHisArgValProCys	1811
QY	627	CCTCGGAAGAAAGACTCCAGCCCATGTTGTCAGCTGAAAGAGGAGCGCCAGATG	686
Db	1812	ThrGlyGluLeuGlnArgProLeuMetValSerGlnIle-----	1825
QY	687	CTCTGCATAGACAACCTGTGGCGGAGGAGCTCAAGAGCGCTCCATCTGCTTCTCAGTAC	746
Db	1826	-----LeuGluAla-----	1828
QY	747	GATGACCAGAGCAGTTTCCCTCAGTCAGAGCTCCCTAAGCCAATGACAACCT-----TTA	800
Db	1829	GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaProGlyLeu	1848
QY	801	GTGGGAAGACTTCTGCGCA-----GTACCAGCAAGTTAATCTCATCACACAG-----	848
Db	1849	ValArgArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGlnProPro	1868
QY	849	-----GTTGATAATGGAGCTCTCCATCAGCTGTCAATGGGCTGCC	890
Db	1869	ValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAla-----GlyGlyTrp	1886
QY	891	TTTCCCTCTGGACCT-----	905
Db	1887	GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg	1906
QY	906	-----GCTCTGCAAGGGCCACCCAAAATAACTCTGCTGGGTACTGTGAC-----TGCTTC	936
Db	1907	ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGluGlyProArg	1926
QY	957	TCCAGCGGGGACTTCTGCAACAGC-----	980
Db	1927	AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGlu	1946
QY	981	-----TGC-----AGTCGCAACACCTCGCCCAT	1004
Db	1947	GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis	1966
QY	1005	GAGCTGAGCGCTTCAAGCCATAAGCGGTCTTGTATAGA---AATCTGAAGCTTC	1061
Db	1967	-----CysValTrpArgCysGlnProGlyCysTyr	1976
QY	1062	CAACCAAAA-----ATGGGAAAGCGCTCTG	1088
Db	1977	CysProProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHisCys	1996
QY	1089	GGAGCTGCTAAACTT-----CGACACAGCAAGGCTGCAACTGTAAGCGC---	1133
Db	1997	SerCysLeuAspLeuLeuThrGlyGlnArgHisHisProGlyAlaArgLeuAlaArgPro	2016
QY	1134	TCAGGCTGCTGAAGAAGTACTGTGAGTGTATGAGGCCAAAATCATGTTCTTCCATT	1193
Db	2017	AspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCysThrAspLeu	2034
QY	1194	TGCAAAATGCTTCTTGC AAAA ACTATGAAGAAAGTCCACAGCAACAAAATGCTGATGAGC	1253
Db	2035	---ProCysProAspCys-----	2039
QY	1254	ACACCCCACTACATGGAGCGCTGGGAGCTTTGAGAGCAGCCATTATTTGCCCCAGCGAAG	1313
Db	2040	-----GlyGlyGlyGlnSerLeuHis-ProCysGlyGlnProCys	2052
QY	1314	TTCTCAGGACCTCCAAAACATGAGAAAATAGGAGCGCTTCTCCTGTATCTCTGGGAA	1373
Db	2052	sProArgSerCysGlnAsp-----LeuSerProGlyse	2063

```
Qy 1374 GTAGTGGAGCCACATGCTGCTGCTGCCCGAGGCTGAGGAGCAGAGCAGGAGCAC 1433
Db 2063 rValCysGlnProGlySerValGlyCysGlnPro-----ThrCysG1 2077
Qy 1434 TGTTCCTCCCAAGCTGGCTGACGATGATGCTGGAGGAGTTTGAAGTGCTGTCGAC 1493
Db 2077 yCysProLeuGlyGlnLeuSerGln-----AspGlyLeuCys----- 2089
Qy 1494 ATTCTCCACATCGAGTTCAGTCCAAAGGGCTGAAATTTAGTAGCTGCAAGCTGGTAA 1553
Db 2090 -----ValProProAlaH1 2094
Qy 1554 AGGGAATGCTGTGGCAAGCCTCAGCCTGGGAATGTCACCCAGGAAGCTGTCGCCA 1613
Db 2094 sCysArgCysGlnTyGlnProGlyAlaMetGlyIleProGlyAsnGlnSerArgSerAl 2114
Qy 1614 GGGAGGAGCAGAGCGCGGCATCATG-----CCAGTCTAGCTGTGAGGCTGTG 1661
Db 2114 aGly--SerArg--PheSerSerTrpGluSerLeuGluProGlyGluVal----- 2129
Qy 1662 AGTGATCTGCATGCTACTGGC-----CAGCCTACTCAAGGTATCTCTAAAG 1706
Db 2130 -----ValThrGlyProCysAspAsnCysThrCysValAlaGlyIleLeuGln 2145
Qy 1707 TGAAGCAGGCA-----GAGCCA-----CCTGGGGA---TGGACA 1739
Db 2146 CysGlnGluValProAspCysProAspProGlyValTrpSerSerTrpGlyProTrpGlu 2165
Qy 1740 CTGCCCTCTGCTCCCTGGGAGCCCTCTGGGAGCTCCCTG----- 1781
Db 2166 AspCysSerValSerCysGly-----GlyGlyGluGlnLeuArgSerArgCysAla 2183
Qy 1782 -----CCTGCATAAAGAGGCTGATTTCTACTTGTGTTATGCTGTTGCTTTCAA 1835
Db 2184 ArgProCysProGlyProAlaArgGlnSerArgThrCys----- 2197
Qy 1836 TTGCTTAGTAGTACTCATTCATCAAGTTATTATGAGCCAGCCTCAAGTTAGAGACTAGGC 1895
Db 2198 -----SerThrGlnVal-----CysArgGluAlaGly 2206
Qy 1896 TCTTCTTCAGTGTGACTC-----TGCCCAATCACATACAGTCAGGTGCGCATCA 1946
Db 2207 CysProAlaGlyArgLeuTyArgGlyCysGln----- 2217
Qy 1947 GGGGTTTTTCCAGCCAGCCTGTGCACAGGAGATATGGAGGGGTCGGGTTAGAGCTG 2006
Db 2218 -----ProGlyGluGlyCysProPheSerCysAlaHis---ValThrGlnGlnVal 2233
Qy 2007 GGTGTTGTTGGATTTTGGGTTTTTTCTTCTGTTATTTCTGTTGAAGTGAGAAAAC 2066
Db 2234 Gly-----CysPheSerGluGlyCysGluGluGly 2243
Qy 2067 TGTCTCTGTGCA-----ACCTTTTCTCATATTACTGTGTCACGTCGCTGCTGAC 2120
Db 2244 Cys-HisCysProGluGlyThrPheGln-HisArgLeuAlaCysValGlnGluCys---P 2262
Qy 2121 CAGTCACAGTGACCTCA 2137
Db 2262 roCysValLeuThrAla 2267

RESULT 12
ID ABG21064
XX ABG21064 standard; Protein; 9222 AA.
XX AC ABG21064;
XX XX
XX 18-FEB-2002 (first entry)
XX DT
XX DE
XX DE
XX KW
XX KW
XX food supplement; medical imaging; diagnostic; genetic disorder.
```

```
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
DR N-PSDB; AAS85251.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID NO 51423; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 9222 AA;
SO
Alignment Scores:
Pred. No.: 0.00168 Length: 9222
Score: 140.50 Matches: 140
Percent Similarity: 30.35% Conservative: 57
Best Local Similarity: 21.57% Mismatches: 194
Query Match: 3.51% Indels: 258
DB: 22 Gaps: 40
US-09-743-237-1 (1-2241) x ABG21064 (1-9222)
Qy 516 GGTAGTGTCCAGCGGCGAGCCCTGGAAGACGACGCTTTCAGGCCCCCTTGGCTCAGGAA 575
Db 2496 GlyThrAlaProAlaAlaPheGlyGlyAlaGluCysGlnGlyProThrMetGluAla 2515
Qy 576 TCCTGTTGCAAGTTC-----CCATCATCCAGGAGGAGGAGGAGGCTCCAGCTGC 626
Db 2516 GluPheCysSerLeuArgProCysProAlaSerTrpGlyCysCysHisArgValProCys 2535
Qy 627 CCTCGGAAGAAAGACTCCAGCCCATGTTGTCAGCTGAAAGAGGCGCCAGATG 686
Db 2536 ThrGlyGluLeuGluGlnArgProLeuMetValSerGlnIle----- 2549
Qy 687 CTCTGCATACAACTGTGGCGGAGGAGGAGCTCAAGCGCTCCATCTGCTTCTCAGTAC 745
```

Db 2550 -----LeuGluAla----- 2552
 QY 747 GATGACAGAGAGTTTCCTCAGTCAGAGCTCCCTAAGCCAAATGACAAC-----TTA 800
 Db 2553 GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaProGlyLeu 2572
 QY 801 GTGGGAAGACTTCTGCCA-----GTACAGCGAAGTTAAATCTCATCACAG----- 848
 Db 2573 ValArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGlnProPro 2592
 QY 849 -----GTTGATATGAGCTCTCCCATCAGCTGCTCAATGGGCTGCC 890
 Db 2593 ValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAla-----GlyGlyTrp 2610
 QY 891 TTTCCCTCTGGACCT----- 905
 Db 2611 GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg 2630
 QY 906 -----GCTCTGCAAGGGCCCAAAATACTCTGTCGGTACTGTGAC---TGCTTC 956
 Db 2631 ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGluGlyProArg 2650
 QY 957 TCCAGGGGGACTCTGCAACAGC----- 980
 Db 2651 AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleLeu 2670
 QY 981 -----TGC-----AGCTGCAACAACCTGGCCAT 1004
 Db 2671 GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspAspLeuValHis 2690
 QY 1005 GAGCTCGAGCGTTCAAGCCATAAAGCGCTGCTGTGATAGA---AATCCTGAAGCTTTC 1061
 Db 2691 -----CysValTipArgCysGlnProGlyCysTyr 2700
 QY 1062 CAACCAAAA----- 2720
 Db 2701 CysProProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHisCys 2720
 QY 1089 GGAGCTGCTAAACTT-----CGACACAGCAAGAGGTGCAACTGTAAGCGC--- 1133
 Db 2721 SerCysLeuAspLeuThrGlyGlnArgHisHisProGlyAlaArgLeuAlaArgPro 2740
 QY 1134 TCAGCTGCTGGAAGAACTACTGAGTCTATGAGGCCAAATAATGATGTCTTCTTCCAPT 1193
 Db 2741 AspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCysThrAspLeu 2758
 QY 1194 TGCATATGATGCTTGCAAAAACATATGAAGAAATGCCAAGACGAAATGCTGATGAGC 1253
 Db 2759 -----ProCysProAspCys----- 2763
 QY 1254 ACACCCCACTACATGGAGCTGGGACTTTGAGAGCAGCATTATTGTGCCCGAGCCAAAG 1313
 Db 2764 -----GlyGlyGlyGlnSerLeuHis-ProCysGlyGlnProCys 2776
 QY 1314 TTCTCAGGACCTCCAAACTAGAAAATAAGCAGCGCTTCTCTGTATCTCTCGGGAA 1373
 Db 2776 sProArgSerCysGlnAsp-----LeuSerProGlySe 2787
 QY 1374 GTAGTGGAGGCCACATGTCCTGCTGCTGGCCAGGGTGAGAACAGCAGAGAGAC 1433
 Db 2787 rValCysGlnProGlySerValCysGlnPro-----ThrCysGln 2801
 QY 1434 TGTTCCTCCCAAGCTTGCTCAGCAGATGATCCTGGAGAGTTTGGAGGTGCTGTGCGCAG 1493
 Db 2801 yCysProLeuGlyGlnLeuSerGln-----AspGlyLeuCys----- 2813
 QY 1494 ATTCTCCACATGAGTTCAAGTCCAAGGGCTGAAAATTGAGTAGCGTGAAGCTGTAA 1553
 Db 2814 -----ValProAlaHis 2818
 QY 1554 AGGGGATCCCTGTGGCAAGCCCTCAGCCCTGGGAATCTCACCAGGAGAGCTGGTGCCA 1613
 Db 2818 sCysArgCysGlnTyrGlnProGlyAlaMetGlyIleProGluAsnGlnSerArgSerAl 2838

QY 1614 GGGAGGAGCAGAGCGCGCATCATG-----CCAGGTGAGCTGTGAGGTCTG 1661
 Db 2838 agly-SerArg---PheSerSerTrpGluSerLeuGluProGlyGluVal----- 2853
 QY 1662 AGTGATCTGCATGGTACTGGC-----CAGCCTACTCAAGGTATCTCTAAAG 1706
 Db 2854 -----ValThrGlyProCysAspAsnCysThrCysValAlaGlyIleLeuGln 2869
 QY 1707 TCAAGCAGCGCA-----GAGCCA-----CCCTGGGGA---TGGACA 1739
 Db 2870 CysGlnGluValProAspCysProAspProGlyValTrpSerSerTrpGlyProTrpGlu 2889
 QY 1740 CTGGCCCTCTCTCTGGGAGCGCTCTGGGACTCCCTG----- 1781
 Db 2890 AspCysSerValSerCysGly-----GlyGlyGluGlnLeuArgSerArgCysAla 2907
 QY 1782 -----CCCTGCATAAAAGAGGGTATTTCTACTTGTGTATGTGTTGCTTTCAAA 1835
 Db 2908 ArgProProCysProGlyProAlaArgGlnSerArgThrCys----- 2921
 QY 1836 TTGCTTAGTAGTACCTCCATTCAAGTTTATGAGCCAGCCTCAAGTTAGAGAGCTAGGC 1895
 Db 2922 -----SerThrGlnVal-----CysArgGluAlaGly 2930
 QY 1896 TCTTCTTTCAGGTGGACTC-----TGCCCAAAATCACATACAGTCAGGTGGCCATCA 1946
 Db 2931 CysProAlaGlyArgLeuTyrArgGluCysGln----- 2941
 QY 1947 GGGGTTTTTCAGGCCAGGCCTGTGACAGAGATATGGAGGGGCTCGGTTAGAGCTG 2006
 Db 2942 -----ProGlyGluGlyCysProPheSerCysAlaHis---ValThrGlnGlnVal 2957
 QY 2007 GGTGTTGTTGGATTTTTCGCTTTTCTCTGATTTTCTGCTGAAGTGAAGAAACT 2066
 Db 2958 Gly-----CysPheSerGluGlyCysGluGly 2967
 QY 2067 TCTCTCTCTGCTCA-----ACCTTTTCTCCATAATTACTGTGCACGGTGCCTGCTGAC 2120
 Db 2968 Cys-HisCysProGluGlyThrPheGln-HisArgLeuAlaCysValGlnGluCys---P 2986
 QY 2121 CAGTCACAGTGCACCTCA 2137
 Db 2986 roCysValLeuThrAla 2991
 RESULT 13
 AAU80189
 ID AAU80189 standard; Protein; 966 AA.
 XX
 AC AAU80189;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human TSPI domain containing protein encoded by cDNA FGI01869.
 XX
 KW TSPI; thrombospondin domain; FGI01869; angiogenesis; vasculogenesis;
 KW human.
 XX
 OS Homo sapiens.
 XX
 PN JP2002085059-A.
 XX
 PD 26-MAR-2002.
 XX
 PF 08-SEP-2000; 2000JP-0273778.
 XX
 PR 08-SEP-2000; 2000JP-0273778.
 XX
 PA (KAZU-) 2H KAZUSA DNA KENKYUSHO.
 PA (YOSH) YOSHITOMI PHARM IND KK.
 XX
 DR WPI; 2002-378268/41.
 DR N-PSDB; ABK50390.

XX TSPI domain-containing polypeptide useful for drug compositions -
 XX Claim 1; Page 30-33; 5lpp; Japanese.
 XX The invention relates to a tspl1 (thrombospondin 1) domain-containing
 CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
 CC encoded by cDNAs designated FG06969 and FG01896. Also included are
 CC proteins that are 50% homologous to the proteins and a polypeptide having
 CC at least one deletion, replacement, addition or insertion of amino acid
 CC in the proteins and having at least 8 repetitions of the tspl1 domain.
 CC The polypeptide can be used in drug compositions particularly
 CC for disorders associated with angiogenesis and vasculogenesis. The
 CC present sequence is the tspl1 domain containing protein encoded by
 CC cDNA FG01896.

XX SQ Sequence 966 AA;

Alignment Scores:

Pred. No.: 0.00138 Length: 966
 Score: 136.50 Matches: 120
 Percent Similarity: 29.83% Conservatives: 53
 Best Local Similarity: 20.69% Mismatches: 204
 Query Match: 3.41% Indels: 203
 DB: 23 Gaps: 29

US-09-743-237-1 (1-2241) x AAU80189 (1-966)

QY 435 TGGTGAGACTCCGATGACACTGCTCCAGGTTCCTCCAGGCGAGCAAGCAAGCAAGCAAGTTC 494
 DB 237 TTPSerTrpCysAspArgSerCysGlyGlyGlnSerLeuArgSerArgSerCysSer 256
 QY 495 GTTCAATCAAGAGACAGCTGTGTGCTCCAGGCGGCGCCCTGAAGCGAGCTTC 554
 DB 257 SerProSerLysAsnGlyGlyAlaProCysAlaGlyGluArgHisGlnAlaArgLeu 276
 QY 555 CAGGCCCTCTGGCTCAGGATCCTGTGCAAGTTCCTCATCCAGGAGGCGAGGAG 614
 DB 277 CysAsnProMetProCysGluAlaGlyCys-----ProAlaGlyMetGluValValThr 294
 QY 615 ---CCCTCCAGCTCCCTCCGAG---AAAGACTCCAGCCCGCCAGTGTGTCACCTG 668
 DB 295 CysAlaAsnArgCysProArgArgCysSerAspLeuGlnGlyIleValCysGln--- 313
 QY 669 AAGGAGGCGCCAGATGCTCTGCTAGACAACTGTGGCGGAGGAGCTCAAGCGCTC 728
 DB 313 ----- 313
 QY 729 CATCTGCTTCTCAGTACGATGACAGAGAGTTCCTCAGTCAGAGCTCCCTAAGCCA 788
 DB 314 -----AspAspGlnValCysGlnLysArgCysArgCysProLys--- 326
 QY 789 ATGACAACTTGTAGTGGGAAGACTCTCCAGTACCAGGAGTAAATCTCATCACACAG 848
 DB 327 -----GlySerLeu-----Glu 330
 QY 849 GTTGATAATGAGCTCTCCATCAGCTGTCAATGGGCTGCCCTTCCTCTGGACCTGCT 908
 DB 331 GlnAspGlyGlyCysValPro----- 337
 QY 909 CTGAAGGGCCACCACAAATAACTCTGTCTGGTACTGTGACTGCTTCTCCAGC----- 962
 DB 338 -----IleGlyHisCysAspCysThrAspAlaGlnGly 348
 QY 963 -----GGGACTTCTGCAACAGCTCCAGCTGC----- 989
 DB 349 HisSerTrpAlaProGlySerGlnHisGlnAspAlaCysAsnAsnCysSerCysGlnAla 368
 QY 990 AACAACTGCGCCATGAGCTCGAGCGCTTCAAGCCCAATAAGGCGTCTCTGATAGAAT 1049
 DB 369 GlyClnLeuSerCysThrAlaGlnProCysProProThrHisCysAlaTrpSerHis 388
 QY 1050 CCTGAAGCTTTCCAACCA---AAAATGGGAAAGCGCGCTCTGGAGCTGCTAAACTTCGA 1106

DB 389 TrpSerAlaTrpSerProCysSerHisSerCysGlyProArgGly-----Gln 404
 QY 1107 CACAGCAAGGGTGCACAACTCTAAGCGCTCAGGC-----TGCCTGAAGAAGTACTGTGAG 1160
 DB 405 GlnSerArgPheArgSerCysThrProGluGlyValIleCysGluAspThrGluCysAla 424
 QY 1161 TGTATGAGCGCAAAATCATGTCTTCCATTTGCAAAATGC---ATTGCTTGCACAAAC 1217
 DB 425 ValProGlnAlaTrpThrLeuTrpSerSerTrpSerAspCysProValSerCysGlyGly 444
 QY 1218 TATGAAGAAAGTCCAGAACGAAATGCTGATGACAGCACCCACCTACTGAGCGCTGG 1277
 DB 445 GlyAsnGlnValArgThrArgAlaCysArgAlaAlaProHisHisAlaGlySerPro--- 463
 QY 1278 GACTTTGAGAGCAGCCATTTATTTGCTCCCGCAGCAAGTTCTCAGGACCTCCAAACTGAGA 1337
 DB 464 -----ProCysLeuGlyProAspThrGlnThr 472
 QY 1338 AAAAATAGCAGCGCC-----TTCTCTGCTATCTCTGGGAAGTAGTGGAGGCCACA 1388
 DB 473 ArgGlnGlnProCysProGlyLeuLeuGluAlaCysSerTrpGlyPro-TripGlyProCys 492
 QY 1389 TGTGCTCTGCTGCTGCGCCAGGAGTGGAGAGCAGAGCAGGAGCTGTCCCAAGCTTG 1448
 DB 492 sSerArgSerCys-GlyProGlyLeuAlaSerArgSerGlySer-Cys---ProCysLeu 510
 QY 1449 GCTGAGCAGATGATCTCTGGAGGAGTTGGAGGCTGCTGCGAGATTTCTCCACATCGAG 1508
 DB 511 -----MetalAlaLysAlaAspProThrCysAsnSerThrPheLeuHisLeuAsp 526
 QY 1509 TTCAAGTCCAAAGGGCTGAAATTTAGTAGCTGCAAGCTGGTAAGGG----- 1557
 DB 527 Thr-----GlnGlyCysThr-SerGlyProCysProGly 537
 QY 1558 -GAATGCTCTGCAAGCC---TCAGCCCTGGGAATCTGCACCGAGGAAGCTGTGCGCCA 1613
 DB 537 uGluCysValTrpSerSerTrpSerTrpThrArgCysSer----- 551
 QY 1614 GGGAGGAGCAGAGGCGCGCATCATGCCAGTCACTGTGAGTCTGAGTGTGATCTCAT 1673
 DB 552 -----CysArg 553
 QY 1674 GGTACTGGCGAGCGCTACTCAAGGTATCTTAAAGTGAAGCAGGAGCAGCCCTGGGGA 1733
 DB 553 gValLeuValGln-----GlnArgTrpArgHisGlnGly-- 564
 QY 1734 TGGACACTGCCCTCTGCTGCTGGGAGGCGCTCTGGGAGCTCCCTGCGCTGCATFAAA 1793
 DB 565 -----ProAlaSerArgGlyAlaArgAlaGlyAlaProCysThrArg---Le 579
 QY 1794 AGAGGTGATTTTCTACTCTTGTATGCTTGTCTTCAAAATTC-----TTAGT 1844
 DB 579 uAspGlyHisPheArg-----ProCysLeuIleSerAsnSerGluAspSerCys 596
 QY 1845 AGTACTCTCAATCAAGTTATTATGAGCAGCAGCTCAAGTTAGAGAGCTAGCTCTCTTCA 1904
 DB 596 sThrProProPheGlu-PheHis-Ala-----CysGlySerProCys 609
 QY 1905 GGTGGACTCTGCCAAATACATACAA----- 1931
 DB 610 AlaGlyLeuCysAlaThrHisLeuSerHisGlnLeuCysGlnAspLeuProProCysGln 629
 QY 1932 -----GTGAGTGGCCATCAGGGGTTTTCAGGCCAG 1964
 DB 630 ProGlyCysThrCysProLysGlyLeuLeuGluGlnAlaGlyGlyCysIleProGlu 649
 QY 1965 GCCTGTGAC-----AGGAGATATGGAGGGGCTCGGGTTAGAGCTG 2006
 DB 650 GlucysAsnSerTrpHisThrSerAlaAlaGlyAlaGlyMetThrLeu 665
 RESULT 14
 AAW87504

ID AAW87504 standard; Protein; 1061 AA.
 AC AAW87504;
 XX
 DT 23-FEB-1999 (first entry)
 XX
 DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.
 XX
 KW Human; N-methyl-D-aspartate receptor; NMDAR2C;
 KW NMDA-activated cation-selective ion channel; glutamate receptor.
 XX
 OS Homo sapiens.
 XX
 PN US5849895-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 20-APR-1994; 94US-0231193.
 XX
 PR 20-APR-1994; 94US-0231193.
 PR 20-APR-1993; 93US-0052449.
 XX
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 XX
 PI Daggett LP, Lu C;
 XX
 DR WPI; 1999-069812/06.
 DR N-PSDB; AAW82910.
 XX
 XX DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
 PT assembly of functional glutamate receptor subunits
 PT
 XX Example 3; Columns 261-268; 203pp; English.
 PS
 CC The present sequence represents a human N-methyl-D-aspartate (NMDA)
 CC receptor subunit (NMDAR). The nucleic acid sequence does not contain the
 CC 860 5'-most nucleotides, has an additional 11 nucleotides (AAV82891)
 CC between nucleotides 1300 and 1301, an additional 24 nucleotides
 CC (AAV82890) inserted between nucleotides 23050 and 2351, as set forth in
 CC AAV82889. The cDNA sequence is derived from clone NMDA24. The NMDAR
 CC subunits contribute to the formation of NMDA-activated cation-selective
 CC ion channels. In addition to being useful for the production of NMDA
 CC receptor subunit proteins, the nucleic acids are also useful as probes to
 CC identify and isolate nucleic acids encoding related receptor subunits.
 CC Functional glutamate receptors can be assembled from several NMDA
 CC receptor subunit proteins of one type (homomeric) or from combinations of
 CC subunit proteins of different types (heteromeric). The present invention
 CC also comprises methods for using such receptor subunits to identify and
 CC characterise compounds which affect the function of such receptors, e.g.
 CC agonists, antagonists and modulators of glutamate receptor function. The
 CC invention also comprises methods for determining whether unknown
 CC protein(s) are functional as NMDA receptor subunits.
 XX
 SQ Sequence 1061 AA;
 Alignment Scores:
 Pred. No.: 0.0043 Length: 1061
 Score: 131.50 Matches: 126
 Percent Similarity: 31.20% Conservative: 45
 Best Local Similarity: 22.9% Mismatches: 155
 Query Match: 3.25% Indels: 223
 DB: 20 Gaps: 30
 US-09-743-237-1 (1-2241) x AAW87504 (1-1061)
 QY 1760 CCCAGGGACAGAGGGCCAGTGTCCATCCCGAGGGTGGCTGCTGCTTGCACCTTAAAG 1701
 DB 182 ProAgtTrpLysSerGly-----ProLeuSerSerTrpArgAlaLeuThrLeuAlaGln 199
 QY 1700 GATACCTTCAGTAGGCTG-----GCCAGTACCAGATCATCATCAGACCTCAGAC 1650
 DB 200 GluAlaValSerProThrProCysProAlaAlaGlyArgAlaThrThrProSerAlaAla 219

QY 1649 CTGACCTGCCCATGATGCGCGCTCTGCT----- 1620
 DB 220 GlyThrTrpProThrProSerSerAlaValArgAspSerAlaSerThrSerSerArg 239
 QY 1619 -----CCTCCCTGGCACCAGCTTCTCT---CGGTCCAGATTCGCCAGGCTGAGGCTTGC 1569
 DB 240 SerTrpProGluTrpSerAsnSer-ProThrThrCysThrTrpProThrAlaSerMetal 259
 QY 1568 CACAGGCATTCCTTTACACGCTTGACGCTTACTCAATTTTTCAGCCCTTGGACCTTGA 1509
 DB 259 aserGlyCysAlaAlaTyrGlyThrAla-----LeuGL 270
 QY 1508 CTCGATGTGGAGAATCTGCACAGGACCTTCCAACTCTCCAGGATCATCTGTCTCAGC 1449
 DB 270 y-ArgCys-----ThrThrSerGlyGlnThrTrpPro---SerAlaProSerP 285
 QY 1448 CAAGCTTGGGGAACAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1389
 DB 285 roSerMetArgAsnAla-----ProArgSerThrSerL 296
 QY 1388 TGTGGCTCTCCACTACTTCCAGGAGATACAGGAGAGGCTGCTCTCTCTCTCTCTCTCT 1329
 DB 296 eutyTrpPro-----LeuTrpArgAlaSerValTrp----- 306
 QY 1328 TGGAGGTCTCTGAGAACTTGGCTGGGACAAATAATGGCTCTCTCTCTCTCTCTCTCT 1269
 DB 307 -----TrpLeu-----AlaAlaMetAlaProSerProp 316
 QY 1268 CATGTAGTGGGTGTGCTCATCAGCAATTTTTCGTTCTGGACTTCTTTCATAGTTTTCGA 1209
 DB 316 roArgProSerTrpSerHisIleAla-----LeuGlnCysGlyCysL 330
 QY 1208 AGCAATGCATTTGCAATGGAAGACACATGATTTTGGCTCTCATACCTCACAGTAGTT 1149
 DB 330 euSerCysAlaSerLeuTrp-----TrpProSerProSerSerCysSerS 345
 QY 1148 CTTACAGGACCTCGAGCGCTTACAGTTGCACCCCTTTTGTCTGTCTGCGAAGTTTAGCAGCTCC 1089
 DB 345 erThrSerAlaLeuSerAlaThrThrArgThr----- 355
 QY 1088 CAGACGGCTTTCCTCCCATTTTGGTGGAAAGCTTCAGGATTTCTATCAAGACACGCCCTT 1029
 DB 356 -----SerPro----- 357
 QY 1028 TATGGCTTTGAAGCGCTCGAGCTCATGGCGCA-----GCTTGTGTCAGCTGCA 981
 DB 358 -----GluAlaArgSerProGlyAlaGlnLeuSerLeuSerAlaSerProCysG 374
 QY 980 GCTGTGTGAGAAAGTCCCGCTCGAGAGCAG-TCACAGTACCCAGACAGATTTATTTGG 922
 DB 374 lyCysCys-----GlyArgTrpSerSerThrThrGlnCysProSerArgThrArgGlyA 392
 QY 921 GTGGCCCTTGCAGA-----GCAGGTCCACAG-----GGAAAGGACGCCCA- 881
 DB 392 laProProAlaArgSerTrpPheTrpSerGlyProSerLeuLeuSerSerSerProA 412
 QY 881 ----- 881
 DB 412 spThrArgProThrTrpProProSerSerLysSerAsnThrSerThrLeuCysArgAlaS 432
 QY 880 -----TTGACAGCTGTGGGAGAG 862
 DB 432 erValThrArgSerPheSerGlyLeuLysIleSerThrHisLeuSerAlaSerAlaArgC 452
 QY 861 CTCCTATTCAACCTGTGTGATGAGATTTAACTTCCTGCTACTGTCGACAGATCTTCCCA 802
 DB 452 ysProThrAla-----AlaArgSerGlyThrSerAlaValT 464
 QY 801 CTAAAGTTCTCATTTGGCTTAGGAGCTCTGACTGAGGAAAGCTCTCTCTGCTGCTCTACT 742
 DB 464 hrThrValThrCys-----ThrProThrTrpSerSerSerT 476
 QY 741 GAGGAAGCAGATGGAGCGCTTTGAGC-----TCCCTCG 709

```

Db 476 hrSerAlaArgTrpArgSerProAlaSerArgTrpAlaLeuArgLeuSerLeuS 496
QY 708 CGCCACAGTTCTCTATCAGAGCATCTGGCGCTCTCTTTCAGCTGACAAATCACCATGG 649
Db 496 erProGlySerTrpMetProSer-SerMetMetLeuLeuSerSer-----ThrThrTrp 513
QY 648 GCCTGGAGTCTTCTCTCCGAGGCGAGCTGGAGGCT-----CCCTCTG 607
Db 514 --GlnAlaArgTrpArgAlaAlaSerTrpSerProLeuGlyLeuAlaArgSerLeuLeu 532
QY 606 CCTCCTGGGATGATGGGAACCTGCAACAGGATCTCAGCCAGAGGGGCT----- 556
Db 533 ProLeuAlaThrAlaSerProCysArgArgTrpProThrGlySerGlyProThrTrpArg 552
QY 555 -----CysCysTrpTrpProTrpGlyTrpProCysTrpSerSerProG 609
QY 459 GCAGTACTTCATCG 446
Db 609 lySerThrTrpSer 613

RESULT 15
AAU80188
ID AAU80188 standard; Protein; 961 AA.
XX
AC AAU80188;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human TSPl domain containing protein encoded by cDNA FGO6969.
XX
KW TSPl; thrombospondin domain; FGO6969; angiogenesis; vasculogenesis;
KW human.
XX
OS Homo sapiens.
XX
PN JP2002085059-A.
XX
PD 26-MAR-2002.
XX
PF 08-SEP-2000; 2000JP-0273778.
XX
PR 08-SEP-2000; 2000JP-0273778.
XX
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
PA (YOSH ) YOSHITOMI PHARM IND KK.
XX
DR WPI: 2002-378268/41.
DR N-PSDB; ABK50389.
XX
PT TSPl domain-containing polypeptide useful for drug compositions -
PS Claim 1; Page 23-26; 51pp; Japanese.
XX
CC The invention relates to a TSPl (thrombospondin 1) domain-containing
CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
CC encoded by cDNAs designated FGO6969 and FGO1896. Also included are
CC proteins that are 50% homologous to the proteins and a polypeptide having
CC at least one deletion, replacement, addition or insertion of amino acid
CC in the proteins and having at least 8 repetitions of the TSPl domain.
CC The polypeptide can be used in drug compositions particularly
CC for disorders associated with angiogenesis and vasculogenesis. The
CC present sequence is the TSPl domain containing protein encoded by

```

```

CC cDNA FGO6969.
XX
SQ Sequence 961 AA;
Alignment Scores:
Pred. No.: 0.0109 Length: 961
Score: 127.00 Matches: 146
Percent Similarity: 30.27% Conservative: 58
Best Local Similarity: 21.66% Mismatches: 206
Query Match: 3.14% Indels: 264
DB: Gaps: 42

US-09-743-237-1 (1-2241) x AAU80188 (1-961)
QY 2159 AGCACCTACCTTCTGGTCTCTGAGTCTGCTGAGTCTGCTGAGCAGCGGACCGTGCAGC 2100
Db 387 SerHisTrpSerAlaTrp-----SerProCysSer 396
QY 2099 AGTAATTATGGAGAAAGTTTGGACAGGAGACAAAGTTTCTCACATTCACAGCAAAATACA 2040
Db 397 HisSerCysGlyProArg---GlyGlnGlnSerArgPheArgSerThrSerGlySer 415
QY 2039 GGAAGAAAAA---CGCAAAAAATCCAAACAAACCCAGCTCTAACCCGACCC----- 1989
Db 416 TrpAlaProGluCysArgGluGlnSerGlnSerGlnProCysProGlnProSerCys 435
QY 1988 CCTCCCAT-----TCTCTGTTCACAGGCGCTGG---CCTGG 1956
Db 436 ProProLeuCysLeuGlnGlyThrArgSerArgThr-LeuGlyAspSerTrpLeuGlnG 455
QY 1955 AAAA-----ACCCCTGATGCCACCTGACTTGATGTGATTTGGGCA 1914
Db 455 yGluCysGlnArgCysSerCysThrProGluGly---ValileCys----- 469
QY 1913 GAGTCCACCTGAAGACAGAGCCTAGCTCTCTAAGTGGAGTGGCTCATATACTTGAAT 1854
Db 470 -----GluAspThrGluCysAlaValProGluAlaTrpThrLeuTrpSer-SerT 486
QY 1853 GGAGGTACTACTAAGCAATTTGAAAGCAACACACATAACAAGTACAAATCACCCCTCT 1794
Db 486 rp-----SerAspCysProValSerCysGlyGlyGlyValArgThrA 502
QY 1793 TTTTATGACAGGAGGAGTCCCGAGAGGCCCTCCCGAGGACAGGAGGCGCAGTGTCGA 1734
Db 502 rgAlaCysArgAlaAlaAlaProHisArgSerProProCysLeuGlyProAspThrG 522
QY 1733 TCCCCAGGTG-----CCTCTGCTGCTTGCACCTTTAGGATACCTTTCAGT 1689
Db 522 lnThrArgGlnGlnProCysProGlyLeuLeuGluAlaCysSerTrp----- 537
QY 1688 AGGTGGCCAGTACCATGCAGATCCTCAGACCTCAGACCTGACCTGCCATGATGCGCG 1629
Db 538 -----GlyProTrpGlyProCysSerArgSerCysGlyProGlyLeuAlaSerArgSerG 556
QY 1628 GCCTCTGCTCTCCCTG-----GGCACCAGCTTCTCCGGT 1593
Db 556 lySerCysProCysLeuMetAlaLysAlaAspProThrCysAsnSerThrPheLeuHis 576
QY 1592 CAGATTCACAGGCTGAGGCTTGCACAGCATTCCTCCCTTTACCAGCTTGCACGCTACTC 1533
Db 576 euAspThrGlnGly-----ProCysAlaGlyIleCysAlaThrHisLeuSerHisGlnLeuCysG 618
QY 1532 AATTTTCAGCCCTTGGACTTGAACATGATGGAGAAATCTGGAGAACTCGCAGCAGCCTTCCAAA 1473
Db 581 ystYrSerGlyPro-----CysProGluAspSerCys-----T 592
QY 1472 CTCCTCCA-----GGATCATCTGCTCAGCCCAAGCTTGGGAACAGT 1431
Db 592 hrProPheGluPheHisAlaCysGlySer----- 602
QY 1430 CTCTGCTGCTCTCTCCTCCCTGGCCAGCAGCAGGCGACATG----- 1387
Db 603 -----ProCysAlaGlyIleCysAlaThrHisLeuSerHisGlnLeuCysG 618

```



```
QY 1386 -----TGGCTCCACTACTTCCAGGAGATACAGAGAGGCGCTGCTATTTTCTCAG 1332
Db 618 lNAspLeuProProCysGlnProGlyCysTyrCysProlys-----GlyL 633
QY 1331 TTTTGGAGTCCCTGAGAACTTGGCTGGGACAAATAATGGCTGCTCTCAAAAGTCCACAGG 1272
Db 633 euLeuGlu-----GlnAlaGlyGlyCysIleProProGluGluC 646
QY 1271 CTCCATGT---AGTGGGGTGTGCTCATCAGCATTTTTCGTTCTG-----GACTTTC 1224
Db 646 ysAsnCysTrpHisThrSerAlaAlaGlyAlaGlyMetThrLeuAlaProGlyAspArgL 666
QY 1223 TTTATAGTTTTTTCAAGCAATGCTTGCAAATGGAAGAACACATGATTTTGGCCTCATA 1164
Db 666 euGlnLeuGlyCysLysGluCysArgArgGlyGluLeu-----HisC 682
QY 1163 GCACCTCAGTAGTTCTTCAGGC-----AGCCTGAGCGCTTACAGT-----T 1122
Db 682 ysThrSerGlnGlyCysGlnGlyLeuLeuProLeuSerGluTrpSerGluTrpSerProC 702
QY 1121 GCACCCCTTTC-----TGTCGCGAAGTTT 1098
Db 702 ysGlyProCysLeuProProSerAlaLeuAlaProAlaSerArgThrAlaLeuGluGluH 722
QY 1097 AGCAGCTC-----CCAGAGCGGCTTCCCATTTTGGTTGGAAAGCTTCAGGATTTCT 1044
Db 722 lstrLeuArgaspProThrGlyLeuSerProThrLeuAlaProLeuLeuAlaSerGluG 742
QY 1043 ATCAAGACAGCGCTTTATGGCTTTGAAGCGCTCGAGCTCAT-----GGCGCA---- 997
Db 742 lNHisArgHisArgLeuCysLeuAspProAlaThrGlyArgProTrpThrGlyAlaProH 762
QY 996 -----GGTTGTTGCAGCT 984
Db 762 lSLeuCysThrAlaProLeuSerGlnGlnArgLeuCysProAspProGlyAlaCysProA 782
QY 983 GCAGCTGT-----TCCAGAAGTCCC----- 964
Db 782 spSerCysGlnTrpSerLeuTrpGlyProTrpSerProCysGlnValProCysSerGlyC 802
QY 963 -----CGCTGGAGACAGTCACAGTACC-----CAGACAGATTATTTT 924
Db 802 lypheArgLeuArgTrpArgGluAlaGluAlaLeuCysGlyGlyCysArgGluProT 822
QY 923 GGCTGGCCCTTCAGAGAGTCCAGAGGAAAGGAGCCCATTTGACAGCTGATGGAG 864
Db 822 rp---AlaGlnaspArgLysLeuGlnArgArgAlaLeuProArgLeuGluLeuArgGlyP 841
QY 863 AGCTCCATTATCAACCTGTGTGATGAGATTAACTTCGCTGTACTGCGCAGAGTCTTCC 804
Db 841 roArgHis----- 843
QY 803 CACTAAAGTTGCTATTGGCTTAGGAGCTCTGACTAGGGAACACTGCTGTGGTCATCGTA 744
Db 844 -----CysIleHisProGlyLeuCysGlnProValProT 855
QY 743 CTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGGCCACACATTTCTCTATGCGAGCAT 684
Db 855 hrGln-----LeuCysArgProL 861
QY 683 CTGGG---GCCCTCCTTTTCAGCTGACAAA-----TCACCATGG- 649
Db 861 euGlyProArgSerValSerAlaGlyThrLeuProProArgLeuProLeuSerProTrpP 881
QY 648 -----GGCTGGAGTCTTTCTTCGAG-----GGCAGCTGAGGCCT---- 613
Db 881 roAlaGlyProGlyTrpAlaLeuCysAlaAspLeuLeuProLeuTrpProProGlnC 901
QY 612 -----CCTCTGCCTCTGGGATGATGGGA----- 589
Db 901 ysGlnCysLeuLeuGlyAlaGlyProGlyProGlyAlaAlaGlyLeuProLysLeuH 921
```

QY 588 --ACTTTGCAACAGGATTCTCTGAGCCAGAGGGGCTTGA 553
Db 921 lSLeuCysGlnArgValProGlyValProThrProGly 933

Search completed: April 21, 2003, 11:28:01
Job time : 136.879 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:21:28 ; Search time 23.4917 Seconds
(without alignments)
5613.620 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 Tactctgtgggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09743237/runat_21042003_111945_440/app_query.fasta_1.4494
-DB-issued_Patents_AA -QFWT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237@cgn_1.1.66 @runat_21042003_111945_440 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	119.5	3.0	1185	4	US-09-041-886-23
C 2	119.5	3.0	2594	4	US-08-718-388-7
C 3	119.5	3.0	5405	4	US-08-718-388-9
C 4	118.5	3.0	3118	4	US-09-579-181-1
C 5	117.5	2.9	2972	4	US-09-579-181-2
C 6	113	2.8	498	2	US-08-660-963-12
C 7	112.5	2.8	323	1	US-07-913-107-2
C 8	112.5	2.8	323	1	US-08-459-201-2
C 9	112.5	2.8	323	1	US-08-281-248-2
C 10	112.5	2.8	323	5	PCT-US93-06421-1
C 11	112	2.8	652	2	US-08-751-305-2
C 12	110	2.7	830	6	5378464
C 13	108.5	2.7	610	3	US-08-365-470-3
C 14	108.5	2.7	610	3	US-09-209-668-19
C 15	108.5	2.7	610	4	US-09-009-490A-89
C 16	108.5	2.7	610	6	5217870-2
C 17	107	2.6	1142	3	US-08-993-118-7
C 18	107	2.6	1142	3	US-08-845-528C-7
C 19	107	2.6	1142	4	US-09-061-709-2
C 20	106	2.6	1564	4	US-09-467-997-1
C 21	104	2.6	1093	3	US-08-545-860D-55
C 22	104	2.6	1093	5	PCT-US94-04496-55
C 23	101	2.5	1251	5	PCT-US95-02251-3
C 24	101	2.5	1252	1	US-08-199-780-3
C 25	101	2.5	1252	2	US-08-316-650-3
C 26	100.5	2.5	1253	3	US-08-479-722B-4
C 27	99	2.4	503	4	US-08-999-774A-10
C 28	98.5	2.4	2414	1	US-08-227-536-2
C 29	98.5	2.5	2414	5	PCT-US95-04682-2
C 30	97.5	2.4	2910	1	US-08-466-033-183
C 31	97.5	2.4	2910	2	US-08-444-733-183
C 32	97.5	2.4	2910	2	US-08-464-134-183
C 33	97.5	2.4	2910	2	US-08-461-361-183
C 34	97.5	2.4	2910	2	US-08-485-910-183
C 35	97.5	2.4	2910	5	PCT-US95-06266-157
C 36	96.5	2.4	426	1	US-07-918-023-2
C 37	96.5	2.4	689	4	US-09-177-249-2
C 38	96.5	2.4	689	4	US-09-061-769A-2
C 39	96.5	2.4	1436	2	US-08-652-971-2
C 40	96.5	2.4	1436	2	US-08-991-258A-2
C 41	96.5	2.4	1436	2	US-08-769-399-2
C 42	96.5	2.4	1436	3	US-08-991-953A-2
C 43	96	2.4	749	4	US-09-562-737-95
C 44	96	2.4	1384	4	US-08-976-255-11
C 45	96	2.4	4302	3	US-08-658-136-5

ALIGNMENTS

RESULT 1

US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LT 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids

Sequence 3, Appl
Sequence 19, Appl
Sequence 89, Appl
Patent No. 5217870
Sequence 7, Appl
Sequence 7, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 55, Appl
Sequence 55, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 4, Appl
Sequence 10, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 183, App
Sequence 183, App
Sequence 183, App
Sequence 183, App
Sequence 157, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 95, Appl
Sequence 11, Appl
Sequence 5, Appl

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-23

Alignment Scores:

Pred. No.: 0.00273 Length: 1185
Score: 119.50 Matches: 125
Percent Similarity: 29.91% Conservative: 44
Best Local Similarity: 22.12% Mismatches: 206
Query Match: 2.95% Indels: 191
DB: 4 Gaps: 24

US-09-743-237-1 (1-2241) x US-09-041-886-23 (1-1185)

```
QY 1991 CCCCTCCCATATCTCTCTGTC-----ACAGCCCTGGCTGGAAACCCCTGA 1944
DB 254 ProProThrProThrProLeuSerValSerSerGlyAlaSerGlyAlaProProThr 273
QY 1943 TGGCCACCTGACTTGTATGTGATTTGGCAGAGTCCACCTGAAGAAGAGCCTAGCTCTCT 1884
DB 274 LysProProThrProVal-GlyGlyGlyAsnLeuProSerAlaProProAla-- 292
QY 1893 AACTTGAGCTGGCTCAATAAATCTGAATGAGTACTACTAAGCAATTTGAAAGCAA 1824
DB 292 ----- 292
QY 1823 CACATAACAAGTAGAAATACCCCTCTTTTATGCGAGGCGAGGAGTCCCGAGAGG 1764
DB 293 -----AsnPheProHisValThrProAsnLeuProProProAla 306
QY 1763 COTCCCCAGGACAGGAGGCGCAGTGTCCATCCCCAGGGTGGCTCTGCTGCTTGCATT 1704
DB 306 aLeuArgProLeuAsnAlaSerAlaSerProProGlyLeuGlyAla----- 322
QY 1703 TAGGATACCTGAGTAGCTGCCAGTACCAGTACAGTACCTACAGCTCAGC----- 1649
DB 323 -----GlnProLeuProGlyHisLeuProSerProTyrAlaMetG1 336
QY 1648 -----TGACCTGGCCATGATCGCGGCTCTGCTCCCTCCCTGGG-- 1610
DB 336 yGlnGlyMetGlyGlyLeuProProGlyProGlyGlyProHisLeuAlaProSerPr 356
QY 1609 -CACCAGTCTCTCGGTGCAGATCCCGAGGCTGAGGCTGCCAGGCAATTCCTTTA 1551
DB 356 OhisSerLeuPro-ProAlaSerSerAlaProAlaProProMetArgPheProTyrS 376
QY 1550 CCAGCTTCACGCTACTCAATTTTCAGCCCTGGACTTGAATCGATGTGAGNAATCG 1491
DB 376 erSer-----SerSerSerSerAla 384
QY 1490 CGACAGGACCTTCCAAACTCTCCAGGATCATCTGCTCAGCCAGCTTGGGGAACAGTG 1431
DB 384 laAlaSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnA 404
QY 1430 CTCCTGCTCTGCTCTCCTCACCCTGGGCCAGGAGCAGCATGTGGCTCCACTACTTC 1371
DB 404 laLeuProSerTyrProHisSerPhePro-----ProProThrSerL 418
QY 1370 CCAGGAGATACAGGAGAGGCTGCCTATTTTCTCAGTTTGGAGTCTCTGAGACTT 1311
DB 418 euSerValSerAsnGlnProProLysTyrThrGlnProSerLeuProSerGlnAlaValT 438
QY 1310 GG-----CTGGGACAAATATGGCTGCT 1287
DB 438 rpSerGlnGlyProProProProProProTyrGlyArgLeuLeuAlaAsnSerAsnAlaH 458
QY 1286 CTCAAAGTCCCGAGCTCCATGTAGTGGGTGTGCTCATCAGCAATTTTCTGCTGAGT 1227
DB 458 lsProGlyProPheProSerThrGlyAlaGlnSerThrAlaHisProProValSerT 478
QY 1226 TTCITCATAGTTTTTGGCAAGCAATTCATTTGCAATGAAAGAAC----- 1183
DB 1183 ----- 1183
```

```
Db 478 hrHisHisHisHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnH 498
QY 1182 -----ACATGATTTTGGCCTCAT-----A 1164
Db 498 isHisGlyAsnSerGlyProProProGlyAlaPheProHisProLeuGluGlyGlys 518
QY 1163 GCACTCAC-----AGTAGTTCTTCAGGCGAGCTGAGCGCTTACA 1125
DB 518 erSerHisHisAlaHisProTyrAlaMetSerProSerLeuGlySerLeuArgProTyrP 538
QY 1124 GTTG-----CACCCCTTCTGCTGTGCGAAGATTAGCAGCTC 1090
DB 538 roProGlyProAlaHisLeuProProHisSerGlnValSerTyrSerGlnAlaGlyP 558
QY 1089 CCAGAGGCGCTTCCCATTTTGGTTGGAAGCTTCAGGATTTCTATCAAGACAGCGCT 1030
DB 558 roAsnGlyPro--ProVal----- 563
QY 1029 TTATGGCTTTGAAGCGCTGAGCTCATGGCGCAGGTTTGTGCAGCTGCAGCTGTGCAGA 970
DB 564 -----SerSerSerSer-----AsnS 569
QY 969 AGTCCCGCTCGAGAAGCAGTCACAGTACCAGACAGAGATTATTTGGGTGGCCCTTGC- 911
DB 569 erSerSerSerThrSerGlnGlySerTyr-----ProCysS 581
QY 910 -----AGAGCAGTCCAGAGGAAAGCGCCCATTTGACAGCTGATGGGA 865
DB 581 erHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPhe----- 596
QY 864 GAGCTCCATTATCAACCTGTGTGATGAGATTAACTTCGTGGTACTGGCGCAGAGTCTTC 805
DB 597 --ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValile 613
QY 804 CCATAAAGTTGTCTATTCGCTTAGGAGCTCTGACTGAGGAAACTGCTCTGGTCATCGT 745
DB 614 -----AlaThrValAlaSerSerProAlaGlyTyrLys 624
QY 744 ACTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGGCCACAGTTGTCTATGCAGACA 685
DB 625 ThrAlaSerProProGly-----ProProProTyrGlyLysArgAlaProSer 640
QY 684 TCTGGGCGCTCCCTTCAGCTGACAAATACCATGGGCTGGAGTCTTTCTTCGAGAGGC 625
DB 641 ProGlyAlaTyrLysThrAla-----ThrProProGlyTyrLysProGlySerProPro 658
QY 624 AGCTGGAGGCTCCTCTCCCTCGGATGATGGAACTTCAACAGGATTCCTGAGCCA 565
DB 659 SerPheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProPro 674
QY 564 GAGGGCTGGAAAGCTGCGCTTTCAGGGCTGCGG-----CCTGGCACACTACAC 514
DB 675 AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuPro 694
QY 513 CTGCT 509
DB 694 roAla 695
```

RESULT 2

US-08-718-388-7
Sequence 7, Application US/08718388
Patent No. 6271362
GENERAL INFORMATION:
APPLICANT: MORIKAWA, MINORU
APPLICANT: HARADA, NAKI
TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA

Db 1802 hrLeuProSerSerTyHisGlyAlaValCysGlyLeuCysGly 1816

RESULT 3

US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9

Alignment Scores:

Pred. No.: 0.00633 Length: 5405
Score: 119.50 Matches: 122
Percent Similarity: 31.8% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
DB: 4 Gaps: 34

US-09-743-237-1 (1-2241) x US-08-718-388-9 (1-5405)

QY 1479 TTCGAATCTCTCAGGATCATCTGCT-----CAGCCAAAGCTTGGGAACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCCTCTCTGCTTCTCCTCCTCCTGCGCCAGCAGGC----- 1396
Db 1421 GluValProAspSerProCysLeuProThrProCysProProGlySerGluAsp 1440
QY 1395 -----AGGCACATGTGGCTCCACTTCTCCAGGAGATACAGGAGAGGCTGCC 1345
Db 1441 CysIleProSerHisLysCysProProGluLeuGluLysTyrglnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGTCTGAGACTTGGCTGGGGACAAATAATGGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCGCCAGCTCCA-----TGTAGTGGG--- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495

QY 1257 -----GTGTGCTCATCAGCATTTTTTCTGTCTGAGCTTCTTCATAGTTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTyrr 1508
QY 1212 TGAAGCAATGC-----ATTTCAAATGGAAGAACAACATGATTTTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValGluProTrpArgThrGluThrPheCys 1528
QY 1170 CCTCATACACTCAGTAGTTCTTCAGCAGCGCTTACAGTTGACCGCTTGGC 1111
Db 1529 ProMetGluCysProProAsnSerHisTyrglnLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TGTGTGCAAGTT-----TAGCAGCTCCAGACGCGCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCATTTTTTGGTGGAAAGCTTCAGGATTTCTATCAAGACACGCTTTATGGCTT 1021
Db 1569 AspSerGlyPheLeuTyrr----- 1574
QY 1020 TGAAGCGCTCGAGCTATGCGCGCAGGTTGTGC-----AGCTGCAGCTGTGTGCAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyrrHis 1589
QY 969 AGTCCCGCTGGAGAAGCAGTCACAGTACCAGACAGAGTATTATTTGGTGGCCCTTGC 910
Db 1590 Asn----- 1590
QY 909 GAGCAGGTCCAGAGGGAAGGAGCCCGCCATTCACAGCTGATGGGAGAGCTCATATCAA 850
Db 1591 -----GlyValTyrrGlu 1595
QY 849 CCTGTGTGATGAGATTTAACTTCGCTGCTACTGCGCAGAAAGTCTTCCCACTAAAGTTGCA 790
Db 1596 Pro-----GluGlnThrValLeuIle----- 1602
QY 789 TTGGCTTAGGGAGCTCTGACTCAGGGAACACTGCTGTGCTCATCTGCTGAGGAGCAGAT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTTGAGTCCCTCCGCGCCACAGTTGT-----CTATGAGCAGCATCT 682
Db 1615 LysGlyMet---ValCysGlnGlnHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTCTTGA-----GCTGACAAATCACCATGGGCTGGAGTCTTCTTCCGAGGCG 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGCGCTCCT-----CTGCTCTCTGGATGATGGGAACCTTGCACAGG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlyGlnGlyValCys----- 1665
QY 576 ATTCTGAGCCAGAGGGCGCTGGAAAGCTGCTTTCAGGCGTCCCGCTG----- 526
Db 1666 LeuProAsnTyrglnAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACACTACCACTGCTTTC-----TTTGATTTTC---AACACTTTTCTGCTT 481
Db 1679 ProHisTyrrHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyrrVa 1698
QY 480 GCTTGTCTGCTGGGAACCTGGCAGTAGTTCATCGAGTCTCACCACAAATG-ATTTCGCC 422
Db 1698 IleAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTAGCCTCAGGATGTAGAGGAGGAGATCAGAGGTTCTCTCAAGTTATAT 362
Db 1714 oPheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyrrVa 1732
QY 361 AGCAATTTCAAGGAATTCCTATATAGCTTAGCAATATG---GTTGAGAAGCAGAATTCAT 305
Db 1732 lArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluIleG 1752
QY 304 CAAAACTCTACATAGGT-----TTAAAGAACTGTCCCTTATCTTACAGAGGGGAGAT 251

QY 1748 -----CCTGTCCTGGG 1759
Db 2623 SerSerThrLeuLysGlyLysThrAsnGlyAlaAspProValProGly 2638
RESULT 6
US-08-660-963-12
; Sequence 12, Application US/08660963
; Patent No. 5852187
; GENERAL INFORMATION:
; APPLICANT: Thorner, Michael O.
; APPLICANT: Gaylinn, Bruce D.
; APPLICANT: Horikawa, Reiko
; APPLICANT: Lyons Jr., Charles E.
; TITLE OF INVENTION: MOLECULAR CLONING OF THE OVINE PITUITARY
; TITLE OF INVENTION: GROWTH HORMONE RELEASING HORMONE RECEPTOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: POPHAM, HAIK, SCHNOBRICH & KAUFMAN, LTD.
; STREET: Metropolitan Square Building, Suite 800, 1450
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,963
; FILING DATE: 12-JUN-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Shaughnessy, Brian P.
; REGISTRATION NUMBER: 32,747
; REFERENCE/DOCKET NUMBER: 18046.036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-824-8000
; TELEFAX: 202-824-8199
; TELEX: 248516
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 498 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-660-963-12
Alignment Scores:
Pred. No.: 0.0079 Length: 498
Score: 113.00 Matches: 113
Percent Similarity: 32.45% Conservative: 46
Best Local Similarity: 23.06% Mismatches: 184
Query Match: 2.79% Indels: 147
DB: 2 Gaps: 25
US-09-743-237-1 (1-2241) x US-08-660-963-12 (1-498)
QY 2084 AAGGTTGGACAGGACAAAGTTTCTCACTTCA-----AGCAGAAATACAGGAAGA 2034
Db 18 ArgIleArgHisGluLeuAlaAlaValThrGlyAspGlyGlyArgSerThrGlyPro 37
QY 2033 AAAAAAGCGAAATATCCAAACAAACCCAGCTCTAACCCGACCCCTCCCATATCTCT 1974
Db 38 ProArgLeuThrGlnAlaAlaValAlaGluLeu-----ProProTrpAlaAla 53
QY 1973 GTCACAGGCTGGCTGGAAAAACCCCTGATGGCCACCTGACTTGATGTGATTTGGGCA 1914
Db 54 GlyCysGlyAlaProAlaSerAlaCysTrpAlaPro---CysGlnSerSerTrpAla 72

QY 1913 GAGTCCACCTGAAGAGAGCGCTAGCTCTCTAACTTGAGGCTGCTCATATAACTTGAAT 1854
Db 73 ThrCysThr-----GlnSerValMetSerSerLeu----- 82
QY 1853 GGAGGTACTACTAAGCAATTTTAAAGCAACACATCAACAAAGTAGAATAACACCTCT 1794
Db 83 -----SerGluArgThrSerLysHisValTyLysLeuLeuLysGly----- 96
QY 1793 TTTTATGACAGGAGGAGTCCCCAGAGGGCTCCCCAGGAGGAGGAGGAGGAGGAGTCCA 1734
Db 97 -----CysPro 98
QY 1733 TCCGAGGCTGGCTC-----TGCTGCTTGCACCTTTAGGATACC 1695
Db 99 ThrProProTrpAlaAlaProGlySerGlyThrGlyCysCysAlaGlyArgTrpGlnAla 118
QY 1694 TTGACTAGGCTGGCCAGTACCATGCAGATCACTCAGACCTCAGAGCTGACCTGGCCATGA 1635
Db 119 LeuGluSerGlyAlaSerProAlaArgLeuSerSerLeuThrSerAlaTrpSerGlnGly 138
QY 1634 TCGCGGCGCTCTGCTCTCCCTGGGACACAGCTTCC---TCGGTGACAGATTTCCAGGGCT 1578
Db 139 LeuArgGlyThrAlaProLeuGlnAlaGlyArgSerProSerArgLeuLeuProArgPro 158
QY 1577 GAGGCTTGCCACAGGCACTTCCCTTTTACCAGCTTGCCAGCTTCACTAAATTTTCAGCCCTT 1518
Db 159 AlaLeuCys-----Pro-TrpSerCysLeuArgArgAsnProThrSerProArg 174
QY 1517 GGACTTGAACATGATGGAGAAATCTGCAGACAGGACCTTCCAAACTCTCCAGGATCAT 1458
Db 174 gGlySerSerThrProTrpAlaThrAlaSerArgLeuGlnProSerTrp-ProSerSerS 194
QY 1457 CTGCTCAGCAAGCTTGGGACAGCTGCTCTGCTCTGCTCTCCTCACCCTGGCCAGCAG 1398
Db 194 erTrpSerLeuSerGlyGlySerThrAlaProGlyThrThrSerThrProSerCysSer- 213
QY 1397 GCAGGCACATGTGGCTCCTCACTACTTCCAGGAGATAC-----AGGAGAAGCCTGCCT 1344
Db 214 -----ProProLeuSerSerArgArgGlnLeuCysSerArgThrProProS 229
QY 1343 ATTTTTCACATTTTGGAGGCTCCTGAGAACTTGGCTG----- 1306
Db 229 erPhetrGly-----ArgThrTrpThrThrAlaAlaSerProLeuSerC 244
QY 1305 -----GGGCAATATATGCTGCTCTCAAA 1281
Db 244 ysAlaArgLeuLeuProProLeuLeuSerArgProProThrSerAlaGlyCysTrpGlnL 264
QY 1280 GT-----CCGAGGCTCCATGTAGTGGGTGTGCTCATCAGCAT 1242
Db 264 ysLeuCysThrProAlaSerProProHisCysProAlaGlnGlyGlySerSerGlyGlyT 284
QY 1241 TTTTCTGCTCTG-----GACTTTCTTCATAGTTTTCAGCAATGCATTTGCAATGGA 1188
Db 284 rpPheSerLeuProGlyGlyPheLeu-----CysSerSerProAlaCysGlyTrpV 301
QY 1187 AGAACACATGATTTGGCCT----- 1168
Db 301 alAlaSer-----trpProLeuLysMetLeuArgAlaGlyThrTrpThrThrAlaProP 319
QY 1167 -----CATAGCACTCAGTAGTCTTCA-----GGCAGCTGAGCGCTTACA 1125
Db 319 roThrGlyGlySerSerLysAspProSerSerSerLeuLeuGlyThrLeuGlyPhePheS 339
QY 1124 GT---TGCACCTTTGCTGTCGAGAGTTTAGCAGCTCCAGACGCGCTTTCCCAATTT 1068
Db 339 erIleLeuSerValSerCysGlyAsnTrpSerGlnLeuArgAlaAlaSerThrProSerL 359
QY 1067 TGGTTGGAAGCTTCAGGATTTCTATCAGACACAGCGCTTTATGGCTTTTGAAGGCTCGAG 1008
Db 359 eu-----SerThrGlyValSerLeu---SerGlnArgP 369
QY 1007 CTCATGGCGCAGGTTGTTGC-----AGTCGACGCTGT----- 976

APPLICANT: Nabi, Ivan R.
 APPLICANT: Otto, Thomas
 APPLICANT: Watanabe, Hideomi
 TITLE OF INVENTION: Method of Determining Metastatic
 TITLE OF INVENTION: Potential of Tumor Cells
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESS: DYKEMA GOSSETT
 STREET: 400 Renaissance Center
 CITY: Detroit
 STATE: MI
 COUNTRY: USA
 ZIP: 48243
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/459,201
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/913,107
 FILING DATE: 14-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Kelly, Robert L.
 REGISTRATION NUMBER: 31,843
 REFERENCE/DOCKET NUMBER: 61,686-016
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (313) 540-0849
 TELEFAX: (313) 540-0763
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 323 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-459-201-2

Alignment Scores:
 Pred. No.: 0.00699 Length: 323
 Score: 112.50 Matches: 86
 Percent Similarity: 36.68% Conservative: 42
 Best Local Similarity: 24.64% Mismatches: 122
 Query Match: 2.81% Indels: 99
 DB: 1 Gaps: 18

US-09-743-237-1 (1-2241) x US-08-459-201-2 (1-323)

Qy 444 TCCGATGAAGTCTCCAGGTTCCCAAGGAGCAAGCAAGCAAGCAAGCAAGTGTCAATC 503
 Db 33 SerGluSerPheLeuPro---SerGluGlyAlaSerSerAspProValThrLeuArg 51
 Qy 504 AAGAAGCAGGTGTGTGTCCAGGCGGAGCCCTGAAGCAGCGAGTTCACGCGCCCT 563
 Db 52 ArgMetLeuAlaAlaArgAsnGlyPheArgSerSerArgProProSerAlaPro 71
 Qy 564 CTGCTCAGGAATCCTGTGTGCAAGTCCCATCATCCAGGAGGAGGAGGCGCTCCAGC 623
 Db 72 Leu-----ProSerSer-----AlaAlaSerCysAlaLeu 81
 Qy 624 TGCCTC-----CGAAGAAAGACTCCAGCCCCATGGTGATTGTCACTGAAGGA 674
 Db 82 CysProThrAspTrpArgArg-----ProValProIleLeuProLeuHisGly 97
 Qy 675 GGCAGCCAGATGCTCTGCATAGACAACTGTGGCGGAGGAGCTCAAGCGCTCCATCTG 734
 Db 98 LysAla-----GlyLeuThrAla 103
 Qy 735 CTTCCTCAGTACGATGACGAGCAGGAGTTCCTCAGTCAGAGCTCCCTAAGCCAAATG--- 791
 Db 104 LeuProLeuTyrLysAlaCysGlyLeuIleValPheGlyGlnLeuIleAsnLeuIleLeu 123

Qy 792 -----ACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCAAGTTAAATCTCATC 842
 Db 124 LeuCysAsnThrPheTyrValThrPheLeuPheProLeuGluThr---LeuGlnIleLeu 142
 Qy 843 ACACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCAATGGGCTGCTTTCCCTCTGGA 902
 Db 143 ThrVal-----GlyMetIleSerSerGlyValAspTrpThrAlaTrpGlyGly 159
 Qy 903 CCTGCTCTGCAAGGGCCACCCAAATAACTGTGTGGGTACTGTGACTGTTCTTCCA-- 960
 Db 160 ArgSer---GlyGlySerGluProValAla-CysLeuGlnGlnAlaAlaSerThrProAl 178
 Qy 961 -----GGGGGACTTCTGCAACAGCTGCAGCTGCAGCTGCACCAACCTG 998
 Db 178 aSerCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
 Qy 999 CGCCATGAGCTCGAGCGCTTCAAGGCCATAAAGCGCTGTCTTGTATAGAAATCTTGAAGCT 1058
 Db 198 lGlyGluAlaHisSerValSerProProArgGlyValThrSerValIleLysLe 218
 Qy 1059 TTCCAACCAAAATGGGAA-----AGGCGTCTGGGAGCTGCT 1097
 Db 218 uLeuSerLeuLeuTrp-LysHisValAspCysAlaArgAlaArgProThrGlySerCys- 237
 Qy 1098 AAAGCTTCGACACAGCAAGGGTGCACACTGTAAGCGCTCAGGCTGCCTGAAGCAACTACTGT 1157
 Db 238 --ThrProGluGlnGlyIle-----LeuGluLysGluLeuLeu- 250
 Qy 1158 GAGTGTATGAGGCCAAATCATGTGTCTTCCATTGTCAAATGCATTGCTTGGCAAAAC 1217
 Db 251 -----ValArgTyrLeuGluGlnA 257
 Qy 1218 TATGAAGAAAGTCCAGACAGAAATGCTGATGACACACACCCCTACATCAGGCTGGG 1277
 Db 257 rgArgGlyLysSerArgAlaIleGlyCysAspGluValThrPro----- 271
 Qy 1278 GACTTTGAGAGCAGCCATTATTGTCGCCAGCCCAAGTTCTCAGGACCTCCAAAAGTGA- 1336
 Db 272 -----PheCysProThr-----ThrSerGlyThrAsp 281
 Qy 1337 -----AAAAATAGCAGCGCTTCTCTGTATCTCTGGGAAGTAGTGGAGGCC 1385
 Db 281 heProSerLeuGlnSerLysAlaGlyLeuIleSerValAsnSerGlyAlaProAlaSerH 301
 Qy 1386 ACATGTGCTCCTCGCTGGGCCA 1408
 Db 301 isGluCysAlaProTrpValPro 308
 RESULT 9
 US-08-281-248-2
 ; Sequence 2, Application US/08281248
 ; Patent No. 5650500
 ; GENERAL INFORMATION:
 ; APPLICANT: Raz, Avraham
 ; APPLICANT: Nabi, Ivan R.
 ; APPLICANT: Otto, Thomas
 ; APPLICANT: Watanabe, Hideomi
 ; TITLE OF INVENTION: Method of Determining Metastatic
 ; TITLE OF INVENTION: Potential of Tumor Cells
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DYKEMA GOSSETT
 ; STREET: 400 Renaissance Center
 ; CITY: Detroit
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48243
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/281,248
;; FILING DATE: 27-JUL-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA: US 07/913,107
;; APPLICATION NUMBER: 31,843
;; FILING DATE: 14-JUL-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kelly, Robert L.
;; REGISTRATION NUMBER: 31,843
;; REFERENCE/DOCKET NUMBER: 61,686-016
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (313) 540-0849
;; TELEFAX: (313) 540-0763
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 323 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-281-248-2

Alignment Scores:
Pred. No.: 0.00699 Length: 323
Score: 112.50 Matches: 86
Percent Similarity: 36.68% Conservative: 42
Best Local Similarity: 24.64% Mismatches: 122
Query Match: 2.81% Indels: 99
DB: 1 Gaps: 18

US-09-743-237-1 (1-2241) x US-08-281-248-2 (1-323)

```
QY 444 TCCGATCACTACTGCCAGTTCCTCCCAAGCAGCAAGCAAGCAAGCAAGCAAGTGTGAAATC 503
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 33 SerGluSerPheLeuPro---SerGluGlyAlaSerSerAspProValThrLeuArgArg 51
QY 504 AAGAAGCAGGTGGTGTAGTGGCAGCGGCGAGCCCTTGAAGCAGCAGCTTTCAGGCGCCCT 563
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 52 ArgMetLeuAlaAlaArgAnGlyGlyPheArgSerSerArgProProSerAlaPro 71
QY 564 CTGGCTCAGAAATCCGTGTGCAAGTTCCTCCATCATCCAGGAGGAGGAGGAGGAGGAGG 623
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 72 Leu-----ProSerSer-----AlaAlaSerCysAlaLeu 81
QY 624 TGCCCT-----CGGAAGAAGAACTCCAGCCCATCGTGTGATTTGTCAGCTGAAAGGA 674
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 82 CysProThrAspTrpArg-----ProValProIleLeuProLeuHisGly 97
QY 675 GCGGCCACAGTCTCTGCATAGACAACTGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 734
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 98 LysAla-----GlyLeuThrAla 103
QY 735 CTTCCTCAGTACGATGACCCAGCAGCAGTTCCTCCAGTACAGAGCTCCCTAAGCAATG--- 791
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 104 LeuProLeuThrLysAlaCysGlyLeuIleValPheGlyClnLeuIleAsnLeuIleLeu 123
QY 792 -----ACAATTTAGTGGGAAGACTCTCCAGTACCAGCGAAGTTAAATCTCATC 842
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 124 LeuCysAsnThrPheThrPheLeuPheLeuPheProLeuGluThr---LeuGlnIleLeu 142
QY 843 ACACAGTTGATATGAGAGCTCTCCATCAGCTGTCAATGGGCGTCCCTTCCCTCTGGA 902
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 143 ThrVal-----GlyMetIleSerSerGlyValAspTrpThrAlaTrpGlyGlyGly 159
QY 903 CTGTCTCTCAAGGCGCCCAAAATAACTGTCTGGTACTGTGTACTGTGTCTCTCCA-- 960
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 160 ArgSer---GlyCysLeuProValAla-CysLeuGlnGlnAlaAlaSerThrProAl 178
QY 961 -----GCGGGGACTTCTGCAACAGCTGAGCTGCAACACACCTG 998
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 178 aSerCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
QY 999 CCCCATGAGCTCGAGGCTTCAAGCCATAAAGGCGTGTCTTGTATAGAAATCCTGAAGCT 1058
```

```
Db 198 LGlyGluAlaHisSerValSerProProArgAGlyValThrSerValIleLysLe 218
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1059 TTCAACCAAAATGGGAA-----AGCCCTCTGGGAGCTGCT 1097
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 218 uLeuSerLeuLeuTrp-LysHisValAspCysAlaArgAlaArgProThrGlySerCys- 237
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1098 AAATTCGACACAGCAAGGGTCAACTGTAAAGCGCTCAGGCTGCCTGAAGAACTACTGT 1157
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 238 --ThrProGluGlnGlnGlyLe-----LeuGluLysGluLeuLeu- 250
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 1158 GAGTGTATGAGGCCAAATCATGTGTCTTCCATTGCAATGCAATGCTTCCAAATAC 1217
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 251 -----ValArgTyrLeuGluGlnA 257
QY 1218 TATGAAGAAGTCCAGAACGAAATGCTGATGAGCACACCCACCTACATGGAGCGCTGGG 1277
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 257 rgArgGlyLysSerArgAlaIleGlyCysAspGluValThrPro----- 271
QY 1278 GACTTTGAGAGCAGCCATTATTTGTCGCCAGCCCAAGTTCTCAGGAGCTCCAAACTGAG- 1336
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 272 -----PheCysProThr-----ThrSerGlyThrAsp 281
QY 1337 -----AAAAATAGCAGCGCTTCTCTCTGTATCTCTCTGGGAAGTACTGGAGGCC 1385
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 281 heProSerLeuGlnSerLysAlaGlyLeuIleSerValAsnSerGlyAlaProAlaSerH 301
QY 1386 ACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1408
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 301 isGluCysAlaProTrpValPro 308
RESULT 10
PCT-US93-06421-1
; Sequence 1, Application PC/TUS9306421
; GENERAL INFORMATION:
; APPLICANT: Raz, Avraham
; APPLICANT: Nabi, Ivan R.
; APPLICANT: Otto, Thomas
; APPLICANT: Watanabe, Hideomi
; TITLE OF INVENTION: Method of Determining Metastatic
; TITLE OF INVENTION: Potential of Tumor Cells
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DYKEMA GOSSETT
; STREET: 400 Renaissance Center
; CITY: Detroit
; STATE: MI
; COUNTRY: USA
; ZIP: 48243
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06421
; FILING DATE: 19930708
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/913,107
; FILING DATE: 14-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kelly, Robert L.
; REGISTRATION NUMBER: 31,843
; REFERENCE/DOCKET NUMBER: 61,686-016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (313) 540-0849
; TELEFAX: (313) 540-0763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 323 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-06421-1

Alignment Scores:
Pred. No.: 0.00699 Length: 323
Score: 112.50 Matches: 86
Percent Similarity: 36.68% Conservative: 42
Best Local Similarity: 24.64% Mismatches: 122
Query Match: 2.81% Indels: 99
DB: 5 Gaps: 18

US-09-743-237-1 (1-2241) x PCT-US93-06421-1 (1-323)

QY 444 TCCGATGAAGTACTGCCAGGTTCCCAAGCAGCAAGCAAGCAAGCAAGTGTGAAATC 503
Db 33 SerGluSerPheLeuPro--SerGluGlyAlaSerSerAspProValThrLeuArgArg 51
QY 504 AAAGAAGCAGGTGGTAGTGTGCCAGCGGCAGCCCTGAAGACGACGCTTTCAGGCCOCT 563
Db 52 ArgMetLeuAlaAlaAlaArgAsnGlyGlyPheArgSerArgProProSerAlaPro 71
QY 564 CTGGCTCAGGATCTGTGTGCAAGTTCCTCATCCAGGAGCGCAGAGGCTCCAGC 623
Db 72 Leu-----ProSerSer-----AlaAlaSerCysAlaLeu 81
QY 624 TGCCTC-----CGGAAGAAAGACTCCAGCCCATGGTGTGTCAGCTGAAAGCA 674
Db 82 CysProThrAspTrpArgArg-----ProValProIleLeuProLeuHisGly 97
QY 675 GCGGCCAGACTCTGCTGATAGACAACCTGTGCGCGAGGAGCTCAAGCGCTCCATCTG 734
Db 98 LysAla-----GlyLeuThrAla 103
QY 735 CTTCTCTAGTACGATGACCAGAGCAGTGTCCCTCAGTCAGAGCTCCCTAAGCCAATG--- 791
Db 104 LeuProLeuTyfLysAlaCysGlyLeuIleValPheGlyGlnLeuIleAsnLeuIleLeu 123
QY 792 -----ACAACTTTAGTGGGAAGACTTCTGCCAGTACCAAGCAAGTTAAATCTCATC 842
Db 124 LeuCysAsnThrPheTyfValThrPheLeuPheProLeuGluThr----LeuGlnIleLeu 142
QY 843 ACACAGGTTGTAATAGGAGCTCTCCCATCAGCTGCTCAATGGGCGCTTTCCTCTGGA 902
Db 143 ThrVal-----GlyMetIleSerSerGlyValAspTrpThrAlaTrpGlyGly 159
QY 903 CTGCTCTGCAAGGCGCCCAAAATACTCTGTGGTACTGTGACTGTCTCTCCA-- 960
Db 160 ArgSer---GlyGlySerGluProValAla-CysLeuGlnGlnAlaAlaSerThrProAl 178
QY 961 -----GCGGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTG 998
Db 178 aserCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
QY 999 CCGCATGAGCTCGAGCGCTTCAAGCCATAAAGCGGTGCTTGTAGAGAATCTCTGAAGCT 1058
Db 198 lGlyGluAlaHisSerValSerProProArgArgGlyValThrSerValIleLysLe 218
QY 1059 TTCCACACCAAAATGGGNA-----AGGCCGCTGGGAGCTGCT 1097
Db 218 uLeuSerLeuLeuTrp-LysHisValAspCysAlaArgAlaArgProThrGlySerCys- 237
QY 1098 AAACCTTCGACACAGCAAGGGTGCAACTGTAAAGCGCTCAGGCTGCTGAAGAACTACTGT 1157
Db 238 --ThrProGlnGlnGlyIle-----LeuGluLysGluLeuLeu- 250
QY 1158 GAGTCTATGAGGCCAAATCATGTGTTCTTCCATTTCGAAATGCTTCTTCAAAAAC 1217
Db 251 -----ValargTyfLeuGluGlnA 257
QY 1218 TATGAAGAAAGTCCAGACGCAAAATGTGTGATGACACACCCCACTACATGAGCGCTGG 1277
Db 257 rgArgGlyLysSerArgAlaIleGlyCysAspGluValThrPro----- 271
```

```
QY 1278 GACTTTGAGAGCAGCCATTATTGTTCCCGCAGCAAGTCTCTCAGGACCTCCAAAACCTGAG- 1336
Db 272 -----PheCysProThr-----ThrSerGlyThrAspP 281
QY 1337 -----AAAAAATAGCAGGCGCTTCTCTGTATCTCTCTGGAAGACTAGTGGAGGCC 1385
Db 281 heProSerLeuGlnSerLysAlaGlyLeuIleSerValAsnSerGlyAlaProAlaSerH 301
QY 1386 ACATGTGCTGCTGCTGCGCCA 1408
Db 301 lSgluCysAlaProTrpValPro 308

RESULT 11
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751.305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Alignment Scores:
Pred. No.: 0.0116 Length: 652
Score: 112.00 Matches: 129
Percent Similarity: 26.88% Conservative: 53
Best Local Similarity: 19.05% Mismatches: 182
Query Match: 2.79% Indels: 313
DB: 2 Gaps: 30

US-09-743-237-1 (1-2241) x US-08-751-305-2 (1-652)

QY 39 GCGTCAGGCACACAGGCGATACACAG-----TGTTGTTCTCTGCTTGGACT 89
Db 19 GlyAlaGlyThrGlyAlaAspThrGluAlaValValCysValGlyThrAlaCysfyrThr 38
QY 90 TGTGACTCCACCCACTCCGCCCGCAGCGGCT----- 122
Db 39 AlaHisSerGlyLysLeuSerAlaAlaGluAlaGlnAsnHisCysAsnGlnHisGlyGly 58
QY 123 -----AGGATAGAACCCAGGCGCTTTTGGCTTCTG-----CAG 158
Db 59 AsnLeuAlaThrValLysSerLysGluAlaGlnHisValGlnArgValLeuAlaGln 78
```

```
QY 159 ATAGTC-----164
Db 79 LeuLeuArgArgGluAlaAlaLeuThrAlaAlaArgMetSerLysPheTrpIleGlyLeuGln 98
QY 165 -----TTCAGCCTGTAGTATTGGGTGGCTGGGA 194
Db 99 ArgGluLysGlyLysCysLeuAspProSerLeuProLeuLysGlyPheSerTrpValGly 118
QY 195 -----GATTTTTCCTTCACACCAAGACTTCCATTATTGAGGATTTTTCAGT 245
Db 119 GlyGlyGluAsp-----ThrProTyrSerAsnTrpHisLysGlu-----131
QY 246 TGATGATCCTCCCTCTGTAAAGACAGTCTCTTAAACCTATGTAGAGTTTGA 305
Db 131 -----131
QY 306 TGAATTCCTGCTTCAACCATPATTGCTAAGCATATATAGCAATTCCTTGAATTCCTATAT 365
Db 132 -----LeuArgAsnSerCysIleSerLys-----ArgCysValSerLeuLeu 146
QY 366 AACTTAGGAGAACCTCTGATT-----CTCCTGCTCTACATCTGCTAGTGTGTAC 419
Db 147 AspLeuSerGlnProLeuLeuProAsnArgLeuProLys-----159
QY 420 AGGGGAAATCATTTTGTGTAGACTCCGATGAACCTACTGCTGAGGTTCCTCAAGGCAGCAAG 479
Db 160 -----TrpSerGluGlyProCysGlySerProGlySerProGlySer---173
QY 480 CAAGCAAGAAAAGTGTGAATCAAGAACAGCAGGTGGTGTGCGCAGGC-----530
Db 174 -----AsnIleGluGlyPheValCysLysPheSerPheLysGlyMetCysArg 189
QY 531 -----GCAGCCCTGAGAGCCAGCCTTTCAGGCCCTCTGCTGCTCAGCAATCC 578
Db 190 ProLeuAlaLeuGlyGlyProGlyGlnValThrTyrThrProPheGlnThrSer 209
QY 579 TGTGTCAGATTCCTCCAGGAGCAGGAGGAGGCTCCAGCTGC-----626
Db 210 -----SerSerLeuGluAlaValProPheAlaSerAlaAlaAsnValAla 224
QY 627 -----CCTCGAAGAAGACTCCAGCCCTAGTGTGTTGCTAGCTGAAA-----671
Db 225 CysGlyGluGlyAspLysAspGluThrGlnSerHisTyrPheLeuCysLysGluLysAla 244
QY 672 -----GGAGGCGCCAGATGCTCTGCTAGACAACTGTGGCGCAGG 713
Db 245 ProAspValPheAspTrpGlySerSerGlyProLeuCysValSer-----259
QY 714 GAGCTCAAAGCGCTCCATCTCTTCTCAGTACGATGACCAGACAGTTCCTCCTCAGTCA 773
Db 260 -----ProLysTyr-----262
QY 774 GAGCTCCCTAAGCCANTGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAAGCAAGTTA 833
Db 262 -----262
QY 834 AATCTCATCACACAGGTGTGATAATGAGCTCTCCCATCAGCTGTCAATGGGCTGCCTTT 893
Db 263 -----GlyCysAsnPhe 266
QY 894 CCTCTGACCTGCTCTGCAAGGCCACCAAAAATAACTGTCTGGGTACTGTGACTGC 953
Db 267 AsnAsnGlyGlyCysHisGln-----AspCys 275
QY 954 TTCTCCAGCGGGGACTTCTGCAACAGCTGCAACACCTGCGCCCATGAGCTCGAG 1013
Db 276 PheGluGlyGlyAspGlySerPheLeuCysGlyCys-----ArgProGlyPheArg 292
QY 1014 CGCTTCAAGCCATAAAGCGCTGTCTCATAGAAATCTGAAGCTTTCACCAACCAAAATG 1073
Db 293 LeuLeuAspAspLeuValThrCysAlaSerArgAsnProCysSerSerProCysArg 312
```

```
QY 1074 GGAAAGGCGCTCTGGGAGCTGTAAACTTCGACACAGCAAGGGTGCAACTGTAAAGCG 1133
Db 313 Gly-----GlyAlaThrCysValLeu 319
QY 1134 TCAGGCTGCCTGAAGAACTAC---TGTGAGTGTCTATGAGCCAAATCATGTGTCTCTCC 1190
Db 320 GlyProHisGlyLysAsnTyrThrCysArgCysProGlnGlyTyrGlnLeuAspSerSer 339
QY 1191 ATTTGCAAAATGCATTGCTTGCAAAACACTATCAAGAAAGTCCAGAACCAAAATCTCTCATG 1250
Db 340 GlnLeuAspCysValAspValAspValAspGluCysGlnAspSerProCysAlaGlnGluCysVal 359
QY 1251 AGCACACCCCACTACATGAGGAGCTTGGGAGCTTTGAGAGCAGCCATTATTGTTCCCAAGCC 1310
Db 360 AsnThrPro-----362
QY 1311 AAGTTCACAGACCTCCAAACTGAGAAAAAATAGGAGGCGCTTCTCTCTGTATCTCTCTGG 1370
Db 363 -----GlyGlyPheArgCysGluCysTrp 370
QY 1371 GAAGTAGTGGAGCCACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
Db 371 ValGlyTyrGlu-ProGlyGlyProGly-----379
QY 1431 CACTGTTCCCAAGCTTGGCTGAGCAGATGATCTGAGGAGGTTTGAAGTGTGCTGCTGCTG 1490
Db 380 -----GluGlyAlaCysGlu 384
QY 1491 CAGATTCTCACATCGAGTTCAAGTCCAGGCGCTGAAATTTAGTAGTGGTGAAGCTGG 1550
Db 384 n-----AspValAspGluCysAlaLeuGlu 392
QY 1551 TAAGGGAATGCTGTGGCAAGCCTCAGCCCTGGGAATCTGCACCCGAGGAGGCTGTGCTG 1610
Db 392 YargSerProCys-----AlaGlnGly---CysThrAsnThrAspGly 406
QY 1611 C-----CAGGAGGAGCAGAGCGCGCATCATGCGGAGGCTGAGCTG 1652
Db 406 rPheHisCysSerCysGluGluGlyTyrValLeuAlaGlyGluAspGlyThr--GlnCys 425
QY 1653 TGAGCTCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1688
Db 426 GlnAspValAspGluCysValGlyProGlyProGlyProLeuCysAspSerLeuCysPheAsn 445
QY 1689 ACTCAAGTATCTTAAAGTGCACAGCAGGAGCCACCCCTGG-----1730
Db 446 ThrGlnGlySerPheHisCys---GlyCysLeuProGlyTrpValLeuAlaProAsnGly 464
QY 1731 ---GGATGGACACTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1772
Db 465 ValSerCysThrMetGlyProValSerLeuGlyProProSerGly 479
```

RESULT 12

```
5378464-2
; Patent No. 5378464
; APPLICANT: MCEVER, RODGER P.
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/320,408
; FILING DATE: 08-MAR-1989
; SEQ ID NO: 2
; LENGTH: 830
5378464-2
```

Alignment Scores:

```
Pred. No.: 0.0214 Length: 830
Score: 110.00 Matches: 128
Percent Similarity: 29.88% Conservative: 68
Best Local Similarity: 19.51% Mismatches: 232
Query Match: 2.74% Indels: 25
Db: Gaps: 35
```

US-09-743-237-1 (1-2241) x 5378464-2 (1-830)

QY 615 GCCTCCAGCTGCTCCGCGAAGAACAGCTCCAGCCCATGGTATTGTCAGCTGAAGGA 674
|||||
Db 259 AlaAlaGlnCysProProLeuLys-----IleProGlu 269
QY 675 GGCCTCCAGATGCTCTGCATAGACAAGTGTGGCGGAGGAGCTCAAAGCGCTCCATCTG 734
|||||
Db 270 ArgGlyAsnMetIleCysLeuMetSerAlaLysAla----- 281
QY 735 CTTCCTCAGTACGATGACAGAGAGCTTTCCCTCAGTCAGAG-----CTC 779
|||||
Db 282 ---ProGlnTrpGlnSerSerCysPheSerCysGluGluGlyPheAlaLeuValGly 300
QY 780 CTAAGGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAATTAATCTC 839
|||||
Db 301 ProGluValValGlnCysThrAlaSerGlyValTrpThrAlaProAlaProValCysLys 320
QY 840 ATCACAGAGTTGATAATGGAGCTCTCCCATCA-----GCT 875
|||||
Db 321 AlaValGlnCysGlnHisLeuGluAlaProSerGluGluGlyThrMetAspCysValHisPro 340
QY 876 GTCAATGGGGTGGCTTTCCCTCTGGACCTGCTCTGCAAGGCGCCACCAATAAATCTG 935
|||||
Db 341 LeuThrAlaPheAlaTyrGlySerSerCysLysPheGluCysGlnProGlyTyrArgVal 360
QY 936 TCTGGG-----TACTGTGACTGCTTCTCCAGCGGGAGCTTC----- 971
|||||
Db 361 ArgGlyLeuAspMetLeuArgCysIleAspSerGlyHisTrpSerAlaProLeuProIle 380
QY 972 TGCACAGCTGCAGCTGCAACAACCTGCGCCAT----- 1004
|||||
Db 381 CysGluAlaIleSerCysGluProLeuGluSerProValMetGlySerMetAspCys 400
QY 1005 ---GAGCTGAGCGCTTCAAGCCATATAAAGCGGTGCTGTATAGAAATCTCGAAGCT 1058
|||||
Db 401 SerProSerLeuArgAlaPheGlnTyrAspThrAsnCysSerPheArgCysAlaGluGly 420
QY 1059 TTCCAACCAAAATGGGAAGGCGCTGGGAGCTGCTAAACTCGACAGCAAGG 1118
|||||
Db 421 PheMet-----LeuArgGlyAlaAspIle 428
QY 1119 TGCAACTGTAAAGCTCAGGC-----TGCTGAAGAACTAC 1154
|||||
Db 429 ValArgCysAspAsnLeuGlyGlnTrpThrAlaProAlaProValCysGlnAlaLeuGln 448
QY 1155 TGTGAG----- 1160
|||||
Db 449 CysGlnAspLeuProValProAsnGluAlaArgValAsnCysSerMetProPheGlyAla 468
QY 1161 ---TGCTATGAGCGCAAAATCATGTGTTCTTCC 1190
|||||
Db 469 PheArgTyrGlnSerValCysSerPheThrCysAsnGluGlyLeuLeuValGlySer 488
QY 1191 ATTTGCAAAATGCAATGCTTGCAAAACATATGAAGAA---AGTCCAGACGAAAAATGCTG 1247
|||||
Db 489 ValLeuGlnCysLeuAlaThrGlyAsnTrpAsnSerValProProGluCysGlnAlaIle 508
QY 1248 ATGAGCACACCCACTACATGAGCGCTGGGAG-----TTT 1283
|||||
Db 509 ProCysThrPro---LeuLeuSerProGlnAsnGlyThrMetThrCysValGlnProLeu 527
QY 1284 GAGAGAGCCATTATTTGCTCCCGAGCAAGTTC-----TCA 1319
|||||
Db 528 GlySerSerTyrLysSerThrCysGlnPheIleCysAspGluGlyTyrSerLeuSer 547
QY 1320 GGACCTCAAAACTGAGAAAAAATAGCAGCGCTTCTCTGCTATCTCTCGGAAAGTAGTG 1379
|||||
Db 548 GlyProGluArgLeuAspCysThrArgSerGlyArgTrpThrAspSerProPrometCys 567
QY 1380 GAGGCCACA---TGTGCTGCTGCTGGCCAGGCTGAGGAAGCAGCAGCAGGAGCTGT 1436
|||||

Db 568 GluAlaThrLysCysProGluLeuPheAla-----ProGluGlnGlySerLeuSerCys 585
QY 1437 TCCCAAGCTGGCTGGTACGACATGATCTGGAGGAGTTTGGAAAGCTGCTGTCAGATT 1496
|||||
Db 586 SerAspThrArgGlyGlu-----PheAsnValGlySerThrCysMetPheSerCys 602
QY 1497 CTCCACATCGAGTTCAAGTCCAAAGGGGCTGAAAATTGATAGCTGCAAGCTGTTAAAGG 1556
|||||
Db 603 ---AsnAsnGlyPheLysLeuGlu----- 609
QY 1557 GGAATGCTGTGGCAAGCCTCAGCCCTGGGAATCTGCAC-----CGAGGAAGCTGG 1607
|||||
Db 610 ---GlyProAsnAsnValGluCysThrMetTyrCysArgHisPro 622
QY 1608 ---TGCCAGGAGGAGGAGCGCGCATCATCGCCAGCTGACCTGACCTGAGGCTG 1661
|||||
Db 623 SerAlaThrProProThrCysLysGlyIleAlaSerLeuProThrPro-----GlyLeu 640
QY 1662 AGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1721
|||||
Db 641 ---GlnCysProAlaLeuThrThrProGlyGlnGlyThrMetTyrCysArgHisPro 659
QY 1722 CCACCCCTGGGATGGACACTGGCCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1781
|||||
Db 660 GlyThrPheGly----- 663
QY 1782 CCCTGCATAAAAGAGGGGTGATTTCTACTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1841
|||||
Db 664 ---PheAsnThrThrCystyr----- 669
QY 1842 AGTAGTACCTCCATTCAAGTTATTATGAGCCAGCTCAAGTTAGAGAGCTAGGCTCTTCT 1901
|||||
Db 670 ---PheGlyCysAsn 673
QY 1902 TCAGGTGA-----CTCTGCCAAATCACATACAGTCAGTGCCCATCAGGGGTT 1952
|||||
Db 674 AlaGlyGlyPheThrLeuIleGlyAspSerThrLeuSerCysArg---ProSerGlyGln 692
QY 1953 TTTCAGGC-----CAGGCCTGT-----GACAGGAGATATGGGAGG 1988
|||||
Db 693 TrpThrAlaValThrProAlaCysArgAlaValLysCysSerGluLeuGluValAsnLys 712
QY 1989 GGGTGGGTTAGAGCTG---GGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2045
|||||
Db 713 ProIleAlaMetAsnCysSerAsnLeuTrpGlyPhePheSerTyrGlySerIleCys--- 731
QY 2046 TCTGCTGAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2099
|||||
Db 732 ---SerPheHisCysLeuGluGlyGlnLeuLeuAsnGlySerAlaGlnThr 747
QY 2100 GCTGCAGGTCG-----CCTGCTGACCACTGACAGTGACCT 2135
|||||
Db 748 AlaCysGlnGluAsnGlyHisTrpSerThrValProThrCysGlnAlaGly---Pro 766
QY 2136 CAGACACAGAGGTAGGCTGCTTATTAT-----GCCACACATTTGTTGTTGTTGTTGTTGTTGTTGTT 2186
|||||
Db 767 LeuThrIleGlnGluAlaLeuThrTyrPheGlyGlyAlaValSerThrIleGlyLeuIle 786
QY 2187 GTGAGAATAAACCTTTCAGACTCCCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 2234
|||||
Db 787 MetGlyGlyThrLeuLeuAlaLeuLeuArgLysArgPheArgGlnLys 802

RESULT 13
US-08-365-470-3
; Sequence 3, Application US/08365470
; Patent No. 5632991
; GENERAL INFORMATION:
; APPLICANT: Gimbrone, Jr., Michael A.
; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX

Db 529 erAlaAlaArgThrCysGlyAla-ThrGlyHisTrpSerGlyLeuLeuProThr---Cys 547
Qy 1783 CTGCATAAAAGAGGGTATTTCTACTCTGTGTGTATGTTCTTTCATCA-----AAT 1836
Db 548 GluAlaProThrGluSerAsnIleProLeuValAlaGlyLeuSerAlaAlaGlyLeuSer 567
Qy 1837 TGCTTAGTACCTCCATTCACAGTTATATGAGCCAGCCTCAAGTTAGAGAGCTAGGCT 1896
Db 568 LeuLeuThrLeuAlaProPheLeuLeuTrpLeuArgLysCysLeuArgLysAlaLysLys 587
Qy 1897 CTCTTCAGGTGGACTCTGCCCAATACATACAGTCCAGTGGCCATCAG 1947
Db 588 PheValProAlaSerSerCysGlnSerLeuGluSerAspGlySerTyrGln 604
RESULT 14
US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

Alignment Scores:
Pred. No.: 0.0257 Length: 610
Score: 108.50 Matches: 122
Percent Similarity: 31.09% Conservative: 58
Best Local Similarity: 21.07% Mismatches: 199
Query Match: 2.71% Indels: 200
DB: 3 Gaps: 31

US-09-743-237-1 (1-2241) x US-09-209-668-19 (1-610)

Qy 502 TCAAGAACGACGAGTGTAGTGTGCGAGCGGCAGCCCTGAAGACGACGCTTTCAGGCCC 561
Db 131 SerLysLysLysLeuAlaLeuCysTyrThrAlaAla----- 142
Qy 562 CTCTGGCTCAGGAATCCTGTGTCAGTTCCTCATCCAGGAGCAGAGGCGCTCCA 621
Db 143 -----CysThrAsnThrSerCysSerGlyHis--GlyGluCysValGluThrIleAsnA 160
Qy 622 GCTGCCCTCGGAAGAACACTCCAGCCCATGGTGATTTGTCAGCTGAAAGGCGGCC 681
Db 160 snTyrThrCysLysCysasp-----ProGlyPheSerG 171
Qy 682 AGATGCTCTGC-----ATAGACAACCTGTGGCGGAGGAG-----C 717
Db 171 LyLeuLysCysGluGlnIleValAsnCysThrAlaLeuGluSerProGluHisGlySerL 191
Qy 718 TCAAGCCCTCCATCTGCTTCTCAGTACGATGACCCAGC-----AGTTTCC 765
Db 191 euValCysSerHisProLeuGlyAsnPheSerTyrAsnSerSerCysSerIleSerCysA 211
Qy 766 CTCAGTCAGAGCTCCCTAAGCCAATGACAACTTTA-----CTGGGAAGA---C 810
Db 211 spArgGlyTyrLeuProSerSerMetGluThrMetGlnCysMetSerSerGlyGluTrps 231
Qy 811 TTCTGCCAGTACCAGC-----AAGTTAAATCTCATCACAGGTTGATAATG 858
Db 231 erAlaProIleProAlaCysAsnValValGluCysAspAlaValThrAsnProAlaAsnG 251
Qy 859 GAGCTCTCCCATCAGCTGTCAATGGGCTGCCTTTCCC----- 896
Db 251 LyPheValGluCysPheGlnAsnProGlySerPheProTrpAsnThrThrCysThrPheA 271
Qy 897 -----TCGTGACCTCTCTGCNAGGGCCACCCAAAATACTCTGTCTGGGTACTGTG 948
Db 271 spCysGluGluGlyPheGluLeuMetGlyAlaGlnSerLeu-----G 285
Qy 949 ACTGCTTCTCCAGGGGGACTTC-----TGCAACAGCTGCAGCTGCA 990
Db 285 InCysThrSerSerGlyAsnTrpAspAsnGluLysProThrCysLysAlaValThrCysA 305
Qy 991 ACAACCTGCGCATGAGCTCGAGCGCTTCAAGGCCATAAAGCGGTGTCTGTATAGAAATC 1050
Db 305 rgAlaValArg----- 308
Qy 1051 CTGAAGCTTTCACACCAAAAAATGGGAAAGCCGCTCTCGGA-----GCTGCTAAAC 1101
Db 309 -----GlnProGlnAsnGlySerValArgCysSerHisSerProAlaGlyGluP 325
Qy 1102 TTCACACAGCAAAAGGTGCAAC-----TGTAAGCGCTCAGGCTGCTGAAGAACTACT 1155
Db 325 heThrPheLysSerSerCysAsnPheThrCysGluGluGlyPheMetLeuGlnGly---- 343
Qy 1156 GTGAGTGCTATGAGGCCAAATCATGTGTTCTTCC----- 1190
Db 344 -----ProAlaGlnValGluCysThrThrGlnGlyGlnTrpThrGlnGlnIleP 360
Qy 1191 --ATTTCGAATGATGCTTTCGCAAAAACTATGAAGAAAGTCCAGAACGAAAAATGCTGA 1248
Db 360 roValCysGluAlaPheGlnCysThrAlaLeu---SerAsnProGluArg----- 375
Qy 1249 TGAGCACACCCCACTACATGAG-----CCTGGGAGCTTTGAGAGCAGCCATTATTTGT 1302
Db 376 -----GlyTyrMetAsnCysLeuProSerAlaSerGlySerPheArgTyrGlyS 392
Qy 1303 CCCAGCAAGTTCTCAGGACCTCCA-----AAACTGAGAAAAAATAGGACAG 1350
Db 392 erSerCysGluPheSerCysGluGlnGlyPheValLeuLysGlySerLysArgLeuGlnC 412
Qy 1351 CCTTCTCCTGTATCTCTCGGAAGTAGTGGAGGCCACATGTCCTGCTGCTGGGCCAG- 1409
Db 412 ysGlyProThrGlyGluTrpAspAsnGluLysProThrCysGluAlaValArgCysAspA 432
Qy 1410 -----GTTGAGGAAG 1419
Db 432 laValHisGlnProLysGlyLeuValArgCysAlaHisSerProIleGlyGluPheF 452
Qy 1420 CAGACGAGGACACTGTTCCCAAGTTGGCTGAGCAGATCATCTGAGGAGTTTGA- 1478
Db 452 hrTyrLysSerSerCysAlaPheSerCysGluGluGly-----PheGluLeuTyrGlyS 470
Qy 1479 -----AGTGCCTGTGCGCAGATTCTCCACATCGATTCAAGTCCAAGGGGCTGA 1527
Db 470 erThrGlnLeuGluCysThrSerGln----- 478
Qy 1528 AAATTGAGTAGCGTCAAGCTGGTAAAGGGGAATCCCTGTGGCAAGCCTCAGCCCTGGGA 1587
Db 478 ----- 478
Qy 1588 ATCTGCACCGAGGAAGCTGG-----TGCCCGAGGAGGACGAGGCC 1629
Db 479 -----GlyGlnTrpThrGluGluValProSerCysGlnValValLysCysSerS 495
Qy 1630 GCGCATATGCGCCAGCTCAGCTGTGAGGCTCTCAGTGATCTCATGTACTGCGCAGCCTA 1689
Db 495 erLeuAlaValProGlyLysIle-----AsnMetSerCysSerGlyGluProV 511
Qy 1690 CTCAGGTATCTAAAGTCAAGCAGGAGCCAGCCCTGGGGATGGACACTG----- 1742
Db 511 alPheGlyThrVal---CysLysPheAlaCysProGlu---GlyTrpThrLeuAsnGlyS 529
Qy 1743 -----GCCCTCTGTCTCCCTGGGAGGCCCTCTCGGGACTCCCTGCG 1782
Db 1782 ----- 1782

Db 529 erAlaAlaArgThrCysGlyAla-ThrGlyHisTrpSerGlyLeuLeuProThr---Cys 547
 QY 1783 CCTGCATAAAAGAGGTGATTTTCTACTGTTGTTGTTATGTTGCTTTCA-----AAT 1836
 Db 548 GluAlaProThrGluSerAsnIleProLeuValAlaGlyLeuSerAlaAlaGlyLeuSer 567
 QY 1837 TGTCTAGTAGTACTCATTCAAGTATATATGAGCCAGCTCAAGTTAGAGAGCTAGGCT 1896
 Db 568 LeuLeuThrLeuAlaProPheLeuLeuTrpLeuArgLysCysLeuArgLysAlaLysLys 587
 QY 1897 CTCTTCAGGTGGACTGTCGCCAAATCACATACAAAGTCAGTGGCCATCAG 1947
 Db 588 PheValProAlaSerCysGlnSerLeuGluSerAspGlySerTyrGln 604

RESULT 15

US-09-009-490A-89
 ; Sequence 89, Application US/09009490A
 ; Patent No. 6300491
 ; GENERAL INFORMATION:
 ; APPLICANT: Bennett and Mirabelli
 ; TITLE OF INVENTION: Oligonucleotide Modulation
 ; TITLE OF INVENTION: of Cell Adhesion
 ; NUMBER OF SEQUENCES: 95
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Office of Jane Massey Licata
 ; STREET: 66 East Main Street
 ; CITY: Marlton
 ; STATE: NJ
 ; COUNTRY: USA
 ; ZIP: 08053

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: WORDPERFECT 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/009,490A
 ; FILING DATE: January 20, 1998
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 440,740
 ; FILING DATE: May 12, 1995

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 063,167
 ; FILING DATE: May 17, 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 969,151
 ; FILING DATE: February 10, 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 007,997
 ; FILING DATE: January 20, 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 939,855
 ; FILING DATE: September 2, 1992

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 567,286
 ; FILING DATE: August 14, 1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jane Massey Licata
 ; REGISTRATION NUMBER: 32,257
 ; REFERENCE/DOCKET NUMBER: ISPH-0268
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (609) 810-1515
 ; TELEFAX: (609) 810-1454
 ; INFORMATION FOR SEQ ID NO: 89:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 610
 TYPE: Amino Acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 ANTI-SENSE: no

US-09-009-490A-89

Alignment Scores:
 Pred. No.: 0.0257 Length: 610
 Score: 108.50 Matches: 122
 Percent Similarity: 31.09% Conservative: 58
 Best Local Similarity: 21.07% Mismatches: 199
 Query Match: 2.71% Indels: 200
 DB: 4 Gaps: 31
 US-09-743-237-1 (1-2241) x US-09-009-490A-89 (1-610)
 QY 502 TCAAGAAGCAGGTGTGTAGTGTCCAGGCGGAGCCCTGAAGACGACGAGTTCACAGGCC 561
 Db 131 SerLysLysLeuAlaLeuAlaLeuCysTyrThrAlaAla----- 142
 QY 562 CTCTGGCTCAGGAATCCTGTTGCAAGTTCCTCCATCATCCAGGAGGAGGAGGAGCCCTCCA 621
 Db 143 -----CysThrAsnThrSerCysSerGlyHis--GlyGluCysValGluThrIleAsnA 160
 QY 622 GCTGCCCTCGGAAGAAGACTCCAGCCCTCATGTTGTTGTCAGCTGAAAGAGGAGGCC 681
 Db 160 snTyrThrCysLysCysAsp-----ProGlyPheSerG 171
 QY 682 AGATGCTCTGC-----ATAGCAACTGTGGCCGAGGGAG-----C 717
 Db 171 lyLeuLysCysGluGlnIleValAsnCysThrAlaLeuGluSerProGluHisGlySerL 191
 QY 718 TCAAGGCGTCCATCTCTCCTCAGTACGATGACCCAGAGC-----AGTTTCC 765
 Db 191 euValCysSerHisProLeuGlyAsnPheSerTyrAsnSerSerCysSerIleSerCysA 211
 QY 766 CTCAGTCAGAGCTCCCTAAGCCCAATGACAACTTTA-----GTGGGAAGA---C 810
 Db 211 spArgGlyTyrLeuProSerSerMetGluThrMetGlnCysMetSerSerGlyGluTrpS 231
 QY 811 TTCTGCCAGTACCAGC-----AAGTTAAATCTCATCACAGGTTGATAATG 858
 Db 231 erAlaProIleProAlaCysAsnValValGluCysAspAlaValThrAsnProAlaAsnG 251
 QY 859 GAGCTCTCCCATCAGCTGTCAATGGGCGCTTCC----- 896
 Db 251 lyPheValGluCysPheGlnAsnProGlySerPheTrpAsnThrThrCysThrPheA 271
 QY 897 -----TCTGGACCTGCTCTCAAGGCCACCCCAAAATACTGTCTGGTACTGTG 948
 Db 271 spCysGluGluGlyPheGluLeuMetGlyAlaGlnSerLeu-----G 285
 QY 949 ACTGCTTCTCCAGCGGGGACTTC-----TGCAACAGCTGCAGCTGCA 990
 Db 285 lnCysThrSerGlyAsnTrpAspAsnGluLysProThrCysLysAlaValThrCysA 305
 QY 991 ACAACCTGCGCCATGAGCTCGAGCGCTTCAAGGCCATAAAGCGTCTTGTATAGAAATC 1050
 Db 305 rgAlaValArg----- 308
 QY 1051 CTGAAGCTTTCCAACCAAAATGGGAAAGCGCGTCTGGA-----GCTGCTAAAC 1101
 Db 309 -----GlnProGlnAsnGlySerValArgCysSerHisSerProAlaGlyGluP 325
 QY 1102 TTCGACACAGCAAAAGGTGCAAC-----TGTAAGCGCTCAGGCTGCCCTGAAGAACTACT 1155
 Db 325 heThrPheLysSerSerCysAsnPheThrCysGluGluGlyPheMetLeuGlnGly----- 343
 QY 1156 GTGAGTGTATGAGGCCAAAATCATGTGTCTTCC----- 1190
 Db 344 -----ProAlaGlnValGluCysThrThrGlnGlnTrpThrGlnGlnIleP 360
 QY 1191 --ATTGTCAAATGCATTGCTTGCACAAAATATGAAGAAGTCCAGACGACGAAATGCTGA 1248
 Db 360 roValCysGluAlaPheGlnCysThrAlaLeu---SerAsnProGluArg----- 375
 QY 1249 TGAGCACACCCCACTACATGGAG-----CCTGGGACTTTGAGACACCCCATTTTCT 1302
 Db 376 -----GlyTyrMetAsnCysLeuProSerAlaSerGlySerPheArgTyrGlyS 392

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:28:43 ; Search time 54.3587 Seconds

(without alignments)
6234.536 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tatctgtggttgcccg.....aaaaaaaaaaaaaaaaaa 2241

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 577658

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame_n2p.model -DEV=rlp
-Q=/cgn2_1/USPTO.spool/US09743237/runat_21042003_111945.462/app_query.fasta_1.4494
-DB=PublishedApplications_AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MTN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09743237@cgn_1_1_49@runat_21042003_111945.462
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pcp:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pcp:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	405.5	10.1	438	10	US-09-220-091-13
2	131.5	3.3	1192	9	US-10-189-971-18
3	131.5	3.3	1207	9	US-10-189-971-20
4	131.5	3.3	1477	9	US-10-189-971-8

5	131.5	3.3	1535	9	US-10-189-971-14	Sequence 14, Appl
6	131.5	3.3	1593	9	US-10-189-971-4	Sequence 4, Appl
7	131.5	3.3	4123	9	US-10-213-509-5	Sequence 5, Appl
8	125.5	3.1	1057	9	US-10-189-971-6	Sequence 6, Appl
9	124	3.1	1251	9	US-10-189-971-16	Sequence 16, Appl
10	124	3.1	1342	9	US-10-189-971-24	Sequence 24, Appl
11	124	3.1	1512	9	US-10-189-971-10	Sequence 10, Appl
12	124	3.1	1570	9	US-10-189-971-12	Sequence 12, Appl
13	124	3.1	1628	9	US-10-189-971-2	Sequence 2, Appl
14	119.5	3.0	5405	9	US-10-025-380-1116	Sequence 1116, Ap
15	119.5	3.0	5405	10	US-09-922-217-1116	Sequence 96, Appl
16	116	2.9	652	10	US-09-789-919-96	Sequence 2, Appl
17	113	2.8	349	9	US-09-361-630-2	Sequence 16, Appl
18	112.5	2.8	939	10	US-09-854-845-16	Sequence 6, Appl
19	112.5	2.8	1034	10	US-09-854-845-6	Sequence 8, Appl
20	112.5	2.8	1078	10	US-09-854-845-8	Sequence 12, Appl
21	112.5	2.8	1136	10	US-09-854-845-12	Sequence 5, Appl
22	111.5	2.8	4123	9	US-10-213-509-5	Sequence 20, Appl
23	109.5	2.7	732	9	US-09-977-418-20	Sequence 22, Appl
24	109	2.7	759	9	US-10-189-971-22	Sequence 36, Appl
25	108.5	2.7	610	9	US-09-802-640-36	Sequence 14, Appl
26	107.5	2.7	954	10	US-09-854-845-14	Sequence 2, Appl
27	107.5	2.7	1049	10	US-09-854-845-2	Sequence 4, Appl
28	107.5	2.7	1093	10	US-09-854-845-4	Sequence 10, Appl
29	107.5	2.7	1151	10	US-09-854-845-10	Sequence 40, Appl
30	107	2.6	838	9	US-10-020-079-40	Sequence 38, Appl
31	107	2.6	851	9	US-10-020-079-38	Sequence 36, Appl
32	107	2.6	951	9	US-10-020-079-36	Sequence 34, Appl
33	107	2.6	957	9	US-10-020-079-34	Sequence 7, Appl
34	107	2.6	1142	9	US-10-085-108-7	Sequence 2, Appl
35	107	2.6	1142	10	US-09-899-651-2	Sequence 78, Appl
36	106.5	2.6	1436	9	US-10-042-431-78	Sequence 48, App
37	106.5	2.6	1436	9	US-09-759-1308-448	Sequence 2, Appl
38	106	2.6	610	10	US-09-748-107-2	Sequence 12, Appl
39	105	2.6	547	10	US-09-779-307-12	Sequence 13, Appl
40	105	2.6	547	10	US-09-779-307-13	Sequence 32, Appl
41	103.5	2.6	863	9	US-10-020-079-30	Sequence 30, Appl
42	103.5	2.6	876	9	US-10-020-079-30	Sequence 28, Appl
43	103.5	2.6	976	9	US-10-020-079-28	Sequence 26, Appl
44	103.5	2.6	982	9	US-10-020-079-26	Sequence 4, Appl
45	103.5	2.6	1036	9	US-09-373-967-4	

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Alignment Scores: 2.76e-25 Length: 438
Pred. No.: 405.50 Matches: 132
Score: 44.37%
Percent Similarity: 44.37%
Conservative: 57

Db	381	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	396
QY	777	CTCCCTAAGCCAAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCAGCAAGTTAAAT	836
Db	397	GluProCysAsnLeuCysThrCysLeuGlyGlyPhe-----	408
QY	837	CTCATCACACAGGTTGATTAATAGGAGCTCTCCCATCAGCTGTCATATGGGCGTCCCTTTCCC	896
Db	409	-----ValThrCysGlyArgProCysGluProProGlyCysSerHisPro	424
QY	897	TCTGGACCTGCTCTGCAAGGCCACCCAAATAACTCTGCTGGTACTGT-	947
Db	425	LeuIlePro-----SerGlyHisCysCysProThr	434
QY	948	-----GACTGC-----TTCTCCAGCGGGAC-----	968
Db	435	CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro	454
QY	969	-----TTCTGCAACAGCTGCAGCTGCACAACCTGCGCCATGAGCTCGAGCGCTTC	1019
Db	455	LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln-----	470
QY	1020	AAAGCCATAAGCGCTGCTTGTATGAATCCTGAAGCTTTCCACCAACCAAAATGGGAAA	1079
Db	471	-----AspGlyGluGluPheGluGlyProAlaGlySer	481
QY	1080	-----GGCGCTCTGGGAGCTGCTAACTTCGA-----	1106
Db	482	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	501
QY	1107	-----CACAGCAAGGGTGCAACTGTAAAGCGC---TCAGGCTGC	1142
Db	502	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	521
QY	1143	CTGAAG-----	1148
Db	522	LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys	541
QY	1149	AACTACTGTGAGTGCTATCAG-----GCCAAATCATGTGTCTCTTCATT	1193
Db	542	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	561
QY	1194	TGCAAA-----TGCAATTGCTGCAAAAACATATGAAGAA	1226
Db	562	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	581
QY	1227	AGTCCAGACGAAATAATGCTGATGAGCACACCCCACTACATGGAGCTGGGGAC---TTT	1283
Db	582	GluGlyArgLys-----TyrGluProGlyGluSerPhe	592
QY	1284	GAGAGCAGCCATTATTGTGTCGCCAGCCAAAGTTTC-----TCAGGA	1322
Db	593	GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly	610
QY	1323	CCTCCAAAACCTGAGAAAATAAGCAGGCGCTTCTCTGTATCTCTCGGGAAGTAGTGGAG	1382
Db	611	ProProSerLeuArgCysHisArg-----GlnCysProSer-----LeuValGly	626
QY	1383	GCCACATGTCCTGCTGCTGCGCCAGGTTGAGGAAGCAGACAGGAGCACTGTTCCTCCA	1442
Db	627	CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCysPro	642
QY	1443	AGCTTGGCTGACGAGATGATCC-----	1464
Db	643	ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr	662
QY	1465	-----TGGAGGAGCTTTGGAAGTGC	1484
Db	662	oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl	682
QY	1485	CTCTCCGAGATTCTCCACATCGAGTTCA-----AGTCCAAGGGCGTGAATAATTGAG	1535

Db	582	aCysProGluLeuSerCysProLeuSerGluArgHisThrProProGly-	-----	698
Qy	1536	TAGCGTCAAGCTGGTAAAGGGGAATCCTGTGGCAAGCCTCAGCCCTGGGAATCTGCAC	1595	
Db	699	-SerCysCysProValCysArgGlyCysValValGluAlaGluGlyArgArgValAlaAs	718	
Qy	1596	CGAGGAAGCTGGTCCCGAGGAGGA-----GCAGAGCGCGGCATCATATGCGCCA	1643	
Db	718	pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysHisArgGlyHI	738	
Qy	1644	GGTCAGCTGGA-----GGTCTGAGTGTATCGCATGCTGCTACTGGCCAGCC	1687	
Db	738	sValGlyCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaAlaLys	758	
Qy	1688	TACTCAAGGTATCCTAAAGTGGCAAGCAGCAGCAGCCACCTGG-----	1730	
Db	758	lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG	778	
Qy	1731	-----GGATGGACACTG-----	1742	
Db	778	nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy	798	
Qy	1743	-----GCCCTCTGCTCCCTGGGAGGCC	1765	
Db	798	sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly-	816	
Qy	1766	CTCTGGGACTCCCTCGCTGCATATAAAGAGGGTGATTTCTACTCTGTTGTTATGTGT	1825	
Db	817	---ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro-	830	
Qy	1826	TGCTTTCAAATTCCTAGTAGTACCTCCATTCAAGTTATTATGACCGACCTCAAGTTAG	1885	
Db	831	-----ArgCysLeuProArgProAlaSerCysMetalaPheGlyAspProHisTyrAr	848	
Qy	1886	AGACTAGGCTCT-----TCITCAGCTGGACTGCCCCAATCATACATACAGTCAG	1936	
Db	848	gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLys	868	
Qy	1937	GTGGCCATCAGGGGTTTTCACGCGCAGGCTGTGACAGAGATATGGAGGGGGTGGG	1996	
Db	868	sPysHisSerGlyAspPheSerValHis---ValThrAsnAspArgGlyArgSerG	887	
Qy	1997	GTTAGAGCTGG 2007		
Db	887	lyValAlaTrp 890		
RESULT 3				
US-10-189-971-20				
; Sequence 20, Application US/10189971				
; Publication No. US20030028907A1				
; GENERAL INFORMATION:				
; APPLICANT: Walke, D. Wade				
; APPLICANT: Scoville, John				
; APPLICANT: Turner, C. Alexander Jr.				
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and				
; TITLE OF INVENTION: Same				
; FILE REFERENCE: LEX-0360-USA				
; CURRENT APPLICATION NUMBER: US/10/189,971				
; CURRENT FILING DATE: 2002-07-03				
; PRIOR APPLICATION NUMBER: US 60/302,949				
; PRIOR FILING DATE: 2001-07-03				
; PRIOR APPLICATION NUMBER: US 60/315,634				
; PRIOR FILING DATE: 2001-08-29				
; NUMBER OF SEQ ID NOS: 25				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 20				
; LENGTH: 1207				
; TYPE: PRT				
; ORGANISM: homo sapiens				
US-10-189-971-20				
Alignment Scores:				
Pred. No.:				1207
				0.0228

Alignment Scores:
Pred. No.:

Score:	131.50	Matches:	135
Percent Similarity:	28.72%	Conservative:	56
Best Local Similarity:	20.30%	Mismatches:	215
Query Match:	3.28%	Indels:	260
DB:	9	Gaps:	36

US-09-743-237-1 (1-2241) x US-10-189-971-20 (1-1207)

558	QY	GCCTCTGGCT-----CAGGAATCCTGTTGC-----	584
320	Db	AlaProCysAlaHisProArgGlnGlyProCysProSerCysAspGlyCysLeuTyr	339
585	QY	-----AAGTTCCCATCATCCACGAGGCA-----	608
340	Db	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu	359
609	QY	-----GAGGAGCCCTCCAGCTGCCCTCGGAAGAAAGACTCCAGGCCCATGTGTG	656
360	Db	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys	379
657	QY	ATTGTGCTGAGTGAAGAGCGCCCGAGATGCTCTGCATAGACAACCTGTGGCGCGAGGAG	716
380	Db	ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu	395
717	QY	CTCAAGCGCTCCATCTGCTTCCTCAGTACGATGACCAGAGCAGCTTTCCTCCAGTCAGAG	776
396	Db	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	411
777	QY	CTCCCTAAGCCAATGACAACTTTAGTGGGAGACTTCTGCCAGTACCAGCGAAGTTAAAT	836
412	Db	GluProCysAsnLeuCysThrCysLeuGlyGlyPhe-----	423
837	QY	CTCATCACACAGGTTGATTAATGGAGCTTCGCCATCAGCTGCATGGGGCTGCCTTTCCC	896
424	Db	-----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro	439
897	QY	TCTGGACCTGCTCTGCAAGGGCCACCACCAATAACTTGTCTGGTACTGT-----	947
440	Db	LeuIlePro-----SerGlyHisCysCysProThr	449
948	QY	-----GACTGC-----TTCCTCAGCGGGAC-----	968
450	Db	CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro	469
969	QY	-----TTCTGCAACAGCTGCAGCTGCCACACACCTGCGCCATGAGCTCGAGCGCTC	1019
470	Db	LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln-----	485
1020	QY	AAAGCCATAAAGCGTGTCTGTGATAAAATCTGAAGCTTTTCCACCAAAAAATGGGAAA	1079
486	Db	-----AspGlyGluGluPheGluGlyProAlaGlySer	496
1080	QY	-----GGCCGCTCTGGGAGCTGCTAAATCTCGA-----	1106
497	Db	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	516
1107	QY	-----CACACCAAGGGTGCACTGTAAGCGC-----TCAGGCTGC	1142
517	Db	LeuProCysLysLeuGlnValThrGluArgGlySerCysProArgCysArgGlyCys	536
1143	QY	CTGAAG-----	1148
537	Db	LeuAlaHisGlyGluHisProGluGlySerArgTyrValProProAspSerAlaCys	556
1149	QY	AACTACTGTGAGTGCATGAG-----GCCAAATCATGTGTCTTCCATT	1193
557	Db	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	576
1194	QY	TGCAAA-----TGCATTGCTTGCAAAACATATGAAGAA	1226
577	Db	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	596
1227	QY	AGTCCAGAACGAAAAATGCTGATGACACACCCCACTACATGGAGCCCTGGGAC-----	1283

RESULT 4

US-10-189-971-8

QY	1020	AAAGCCATAAAGCGTCTTGATAGAAAATCCTGAAGACTTTTCACCAAAAAATGGGAAA	1079
Db	756	---:::---AspGlyGluPheGluGlyProAlaGlySer	766
QY	1080	-----GCCCTCTGGAGAGCTGCTAACTTCGA	1106
Db	767	CysLeuTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	786
QY	1107	-----CACACAAGGGTGCAACTGTFAAGCGC---	1142
Db	787	LeuProCysIysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	806
QY	1143	CTGAAG-----	1148
Db	807	LeuAlaHisGlyGluGluHisProGluGlySerArgTyrValProProAspSerAlaCys	826
QY	1149	AAC TACTGTGAGTCTATGAG-----GCCAAANTCATGTGTCTTCCATT	1193
Db	827	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	849
QY	1194	TGCAAA-----TGCAATTGCTTGC AAAA A C T A T G A G A A	1256
Db	847	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspcyGluHis	866
QY	1227	AGTCCAGAAGCAAAAATGCTGATGAGCACACCCTACTACATGGAGCCCTGGGGAC---	1283
Db	867	GluGlyArgLys-----TyrGluProGlyGluSerPhe	877
QY	1284	GAGAGCAGCCATTATTGTGCCCCAGCAAGTTC-----TCAGGA	1322
Db	878	GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly	895
QY	1323	CCTCCAAACTGAGAAAAATAGCAGGCGCTCTCTCTGATCTCTCTGGGAAGTAGTGGAG	1382
Db	896	ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly	911
QY	1383	GCCACATGTCCTGCTGTCGCCCGCCAGGTGAGAGCAGCAGCAGCAGTGTTCOCCA	1442
Db	912	CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCyspro	927
QY	1443	AGCTTGCTCAGCAGATGATCC-----	1464
Db	928	ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr	947
QY	1465	-----TGAGGAGCTTTTGGAAAGGTGC	1484
Db	947	OProAspProCysTyrrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl	967
QY	1485	CTGTGCGAGATTCCACATCGAGTTCA-----AGTCCAAGGGCTCAAAATTGAG	1535
Db	967	aCysProGluLeuSerCysProLeuSerGluArgHisThrProProGly-----	983
QY	1536	TAGCGTCAAGCTGGTAAGGGGAATGCCTGTGGCAAGCCTCAGCCCTCGGANATCTGCAC	1595
Db	984	-SerCysCysProValCysArgGluCysValGluAlaGluGlyArgArgValAlaAs	1003
QY	1596	CGAGGAAGCTGGTCCCGCAGGGAGGA-----GCAGAGCGCGCGCATCATGGCCA	1643
Db	1003	pGlyGluSerTrpArgAspproSerAsnAlaCysIleAlaCysThrCysHisArgGlyHI	1023
QY	1644	GGTCAGCTGTGA-----GGTCTGAGTGCATCTGCATGTGCTACTGCCACGCC	1687
Db	1023	svalGlcCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTyrAlaLysva	1043
QY	1688	TACTCAGGATTCCTTAAGTGCAGCAGCAGCAGCCACCTGG-----	1730
Db	1043	lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysvalHisGl	1063
QY	1731	-----GGATGGACACTG-----	1742
Db	1063	nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy	1083
QY	1743	-----GCCCTCCTGTGCTGGGAGGCC-----	1765

```
Db 1083 smetalaglythrvalargcysglnserclynargcysserproleusercysgly 1101
Qy 1766 CTCCTGGGACCTCCCTGCATGATAAAAGAGGGATTTCTTACTCTGTGTATGTGT 1825
Db 1102 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1115
Qy 1826 TGCTTTCAATTCCTAGTACCTCCTCATTCAAGTATTATGAGCCAGCCCTCAAGTTAG 1885
Db 1116 ----ArgCysLeuProArgProAlaSerCysMetalaPheGlyAspProHisTyrAr 1133
Qy 1886 AGAGCTAGCTCT-----TCTCAGGTGGACTCTGCCAATATCATACACAGTCAG 1936
Db 1133 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1153
Qy 1937 GTGGCCATCAGGGGTTTCCAGGCGAGCCCTGTGCACAGGAGATATGGAGGGGGTCGG 1996
Db 1153 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspArgGlyArgSerG 1172
Qy 1997 GTTAGAGCTGG 2007
Db 1172 lyValAlaTrrp 1175

RESULT 5
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20030028907A1el Human Kielin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-14

Alignment Scores:
Pred. No.: 0.0246 Length: 1535
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
Db: 9 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-14 (1-1535)
Qy 558 GCCCTCTGGCT-----CAGGAATCCTGTTGC----- 584
Db 648 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 667
Qy 585 -----AAGTTCCCATCATCCAGGAGGCA----- 608
Db 668 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 687
Qy 609 -----GAGGAGGCTCAGCTGCCCTCGGAGAAGACTCCAGGCCCATGTGG 656
Db 688 CysLeuCysTrrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 707
Qy 657 ATTTGTACGCTGAAGGAGGCGCCAGATGCTCTGCATACAGAACTGTGGCGCGAGGAG 716
Db 708 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 723
```

```
Qy 717 CTCAAAGCGCTCCATCTCTCTCCTCAGTACGATCACCAGACAGTTCCTCCTCAGTCAGAG 776
Db 724 -----TyrLeuGlyGluSerTyrTyrLeuSerAsnGlnGluPheProAspProArg 739
Qy 777 CTCCTTAGCAANTACAACTTTAGTGGGAAGACTTTCGCCAGTACACGCAAGTTAAAT 836
Db 740 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 751
Qy 837 CTCATCACACAGGTGTGATAATGGAGCTCTCCCATCAGCTGTCAATGGGGCTGCCTTTCCC 896
Db 752 -----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro 767
Qy 897 TCTGGACCTGCTCTGCAAGGCCACCCCAAAATAAATCTCTGTGGTACTGT----- 947
Db 768 LeuIlePro-----SerGlyHisCysCysProThr 777
Qy 948 -----GACTGC-----TTCTCCAGCGGGGAC----- 968
Db 778 CysGlnGlyCysArgTyrHisGlyValThrAlaSerGlyGluThrLeuProAspPro 797
Qy 969 -----TTCTGCAACAGCTGCAGCTGCAACAACCTGCCCATGAGCTCGAGCGCTTC 1019
Db 798 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 813
Qy 1020 AAGCCATAAAGCGCTGTCTTGTAGAAATCCTGAAGCTTTTCCCAACCAAAATGGGAAA 1079
Db 814 -----AspGlyGluGluPheGluGlyProAlaGlySer 824
Qy 1080 -----GGCGTCTGGAGCTGCTAACTCGA----- 1106
Db 825 CysGluTrpCysArgCysGlnAlaGlyValSerCysValArgLeuGlnCysProPro 844
Qy 1107 -----CACAGCAAGGTCGCAACTGTAAGCG-----TCAGGCTGC 1142
Db 845 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 864
Qy 1143 CTGAAG----- 1148
Db 865 LeuAlaHisGlyGluGluHisProGluGlySerArgTrrpValProAspSerAlaCys 884
Qy 1149 AACTACTGTGAGTCTATGAG-----GCCAAATCATGTGTCTTCCATT 1193
Db 885 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 904
Qy 1194 TGCAAA-----TGCAATGCTTGCACAACTATGAAGAA 1226
Db 905 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 924
Qy 1227 AGTCCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGGAGCCTGGGAC---TTT 1283
Db 925 GluGlyArgLys-----TyrGluProGlyGluSerPhe 935
Qy 1284 GAGAGCAGCCATTATTGTCTCCAGCCCAAGTTC-----TCAGGA 1322
Db 936 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 953
Qy 1323 CCTCCAAAACAGCAAAAAATAGCAGCGCTCTCTCTATCTCTCTGGAAGTAGTGGAG 1382
Db 954 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 969
Qy 1383 GCCACATGTGCTGCTGGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1442
Db 970 CysProProSerGlnLeuLeuProGly-----ProGlnHisCysCysPro 985
Qy 1443 AGCTTGGCTGAGCAGATGATCC----- 1464
Db 986 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1005
Qy 1465 -----TGGAGAGTATTGGAGGTGC 1484
Db 1005 oProAspProCysTrrpThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 1025
```

QY 1485 CTGTCACAGATTCTCCACATCGAGTTCA-----AGTCCAAAGGGGCTGAAAATTGAG 1535
Db 1025 aCysProGluLeuSerCysProLeuSerGluArgHisThrProGly----- 1041
QY 1536 TAGCTGCAAGCTGTAAGGGAATGCTGTGGCAAGCTCAGCCCTGGGAATCTGCAC 1595
Db 1042 -SerCysCysProValCysArgGluCysValGluAlaGluGlyArgValAlaAs 1061
QY 1596 CGAGGAAGCTGGTCCACGGAGGA-----GCAGAGGCGCGCATCATGGCCA 1643
Db 1061 pGlyGluSerTrpArgAspProSerAsnAlaCysHisAlaCysThsCysHisArgGlyHI 1081
QY 1644 GGTACCTCTGA-----GGTCTGAGTGTCTGCATGCTGAGTGTGGCCAGCC 1687
Db 1081 sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1101
QY 1688 TACTCAAGTATCCTAAAGTGAAGCAGCAGCAGCCACCTGG----- 1730
Db 1101 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 1121
QY 1731 -----GGATGGACACTG----- 1742
Db 1121 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThsSerCysSerCy 1141
QY 1743 -----GCCCTCTCTCCCTGGGGAGGCC 1765
Db 1141 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 1159
QY 1766 CTCTGGGACTCCCTCCCTGCTGATAAAGAGGGTGAATTTCTACTGTGTGTATGTGTT 1825
Db 1160 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1173
QY 1826 TGCTTTCAAATTGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAG 1885
Db 1174 ----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1191
QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGACTCTGCCCAATACATACAAGTCAG 1936
Db 1191 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1211
QY 1937 GTGGCCATCAGGGTTTTCCAGGCCAGCCCTGTGACAGAGATATGGAGGGGGTCTCG 1996
Db 1211 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 1230
QY 1997 GTTAGAGCTGG 2007
Db 1230 lyValalaTrp 1233

RESULT 6

US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human kielin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Alignment Scores:
Pred. No.: 0.0249 Length: 1593
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
DB: 9 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-4 (1-1593)

QY 558 GCCCTCTCTGGCT-----CAGGAATCCTGTTC----- 584
Db 706 AlaProCysAlaHisProArgGlnGlyProCysProSerCysAspGlyCysLeuTyr 725
QY 585 -----AAGTCCCATCATCCAGGAGCA----- 608
Db 726 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 745
QY 609 -----GAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTG 656
Db 746 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 765
QY 657 ATTTGTCAGCTGAAGAGGCGGCCAGATGCTCTGCATACACAACTGTGGCGGAGGGAG 716
Db 766 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 781
QY 717 CTCAAAGCGCTCCATCTCTCTCAGTAGCAGGACGAGCAGAGCTTCCCTCAGTCAGAG 776
Db 782 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 797
QY 777 CTCCTTAAGCCCAATGAGTGTAGTGGGAAGACTTCTGCCAGTACCGAGCAAGTTAAAT 836
Db 798 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 809
QY 837 CTCATCACACAGGTTGATATGAGAGCTCTCCCATCAGCTGTCAATGGGGTGCCTTTCCC 896
Db 810 -----ValThrCysGlyArgArgProCysGluProGlyCysSerHisPro 825
QY 897 TCTGGACCTGCTCTGCAAGGGCCACCAATAACTCTGCTGGGTACTGT----- 947
Db 826 LeuIlePro-----SerGlyHisCysCysProThr 835
QY 948 -----GACTGC-----TCTCTCAGCGGGGAC----- 968
Db 836 CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro 855
QY 969 -----TCTGCAACAGCTGCAGCTGCACAACTCGGCCCATCAGCTCGAGCGCTTC 1019
Db 856 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 871
QY 1020 AAAGCCATAAAGCGTCTTGATAGAAATCCTGAAGCTTCCCAACCAAAAATGGGAAA 1079
Db 872 -----AspGlyGluGluPheGluGlyProAlaGlySer 882
QY 1080 -----GGCCGTCTGGGAGCTGCTAACTTCCA----- 1106
Db 883 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 902
QY 1107 -----CACAGCAAGGGTGAACGTGAAGCGC-----TCAGGCTGC 1142
Db 903 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 922
QY 1143 CTGAAG----- 1148
Db 923 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 942
QY 1149 AACTACTGTGAGTGTATGAG-----GCCAAAATCATGTCTTCTTCATT 1193
Db 943 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 962
QY 1194 TCGAAA-----TGCATTCTTCAAAAACTATGAAGA 1226
Db 963 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 982

```
QY 1227 AGTCAGAACGAAATGCTGATGAGCACACCCCACTACATGAGCCTGGGAC---TTT 1283
Db 983 GluGlyArgLys-----TyrGluProGlyGluSerPhe 993
QY 1284 GAGACAGCCATTATTTGTCCTCCCAAGTC-----TCAGGA 1322
Db 994 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 1011
QY 1323 CCTCAAACTGAGAAAATAGCAGGCCTTCTCTGTATCTCTCTGGAGTAGTGAG 1382
Db 1012 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 1027
QY 1383 GCCACATGCTGCTGCTGCTGGCCAGGGTGAGGAGCAGCAGGACACTGTTCCCA 1442
Db 1028 CysProProSerGlnLeuLeuProProGly-----ProGlnHisCysCysPro 1043
QY 1443 AGCTTGCTGACGAGATGATCC-----1464
Db 1044 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1063
QY 1465 -----TGGAGGAGTTTGGAGGTGC 1484
Db 1063 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTripleCysIleHisGlnAl 1083
QY 1485 CTGTGCGCAGATCTCCACATCGAGTTCA-----AGTCCAAGGGCTGAAAATTGAG 1535
Db 1083 aCysProGluLeuSerCysProLeuSerGluArgHisThrProProGly-----1099
QY 1536 TAGGTGCAAGCTGGTAAGGGGAATGCTGTGGCAAGCCTCAGCCCTGGAATCTGCAC 1595
Db 1100 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 1119
QY 1596 CGAGAACCTGGTCCCGAGGAGGA-----GCAGAGCCCGGCATCATGCCCA 1643
Db 1119 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHi 1139
QY 1644 GGTGAGCTGTGA-----GCTGAGTGATGCTGCATGGTACTGGCCCAAGCC 1687
Db 1139 sValGluCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1159
QY 1688 TACTCAAGGTATCTAAGTCCAGCAGGAGGAGCCACCCCTGG-----1730
Db 1159 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisGl 1179
QY 1731 -----GGATGGACACTG-----1742
Db 1179 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1199
QY 1743 -----GCCCTCTGCTCCCTGGGAGGCC 1765
Db 1199 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly-----1217
QY 1766 CTCCTGGGACCTCCCTGCTCATAAAGAGGGTGATTTCTACTGTGTTATGTGTT 1825
Db 1218 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro-----1231
QY 1826 TGCTTTCAAATGTGTAGTAGTACTCCATTCAAGTTATTATGAGCAGCCTCAAGTTAG 1885
Db 1232 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1249
QY 1886 AGACTAGGCTCT-----TCTCAGGTGGACTCTGCGCAAAATCACATACAGTCAG 1936
Db 1249 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1269
QY 1937 GTGCCCATCAGGGTCTTTCAGGCCAGGCTGTGCAGAGGAGATATGGAGGGGGTTCGG 1996
Db 1269 spCysHisSerGlyAspPheSerValHis-----ValThrAsnAspAspArgGlyArgSerG 1288
QY 1997 GTTAGAGCTGG 2007
Db 1288 lYValAlaTriP 1291
```

```
RESULT 7
US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 60/311,720
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-213-509-5

Alignment Scores:
Pred. No.: 0.034 Length: 4123
Score: 131.50 Matches: 139
Percent Similarity: 27.79% Conservative: 60
Best Local Similarity: 19.41% Mismatches: 217
Query Match: 3.28% Indels: 300
Db: 9 Gaps: 35

US-09-743-237-1 (1-2241) x US-10-213-509-5 (1-4123)
QY 468 CAAGCGACGCAAGCAAGCAAGAAAGTGTGAATCAAGAAAGCAGGTGGTAGTGCCA 527
Db 3034 GlnGlyProThrMetGluAlaGluPheCysSerLeuArgProCysProGlyProValPro 3053
QY 528 GGGGGAGCCCTGAAGACGACGAGCTTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAG 587
Db 3054 GlyMetCysProArgAspLysGlnTrpLeu-----AspCysAlaGln 3067
QY 588 TTCCATCATCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 647
Db 3068 GlyProLaser-----3071
QY 648 CCCATGCTGATTGCTCAGCTGAAAGAGGAGGCCAGATGCTGTCATAGACAACTGTGGC 707
Db 3071 -----3071
QY 708 GCGAGGAGCTCAAAGCGCTCCATCTGCTTCTCAGTACGATGACGACGAGTTCCTCCT 767
Db 3072 CysAlaGluLeuSerAla-----ProArgGlyThrAsnGlnThrCysHisPro 3087
QY 768 CAGTCAGAGCTCCCTAAGCCAATGACAACCTTTAGTG-----GGAGACTTCTGCCA 818
Db 3088 GlyCysHisCysProSerGlyMetLeuLeuValSerProArgGlyHisProGlyPro 3107
QY 819 GTACCCAGCAAGTTAAAT-----CTCATCACACAGGTTGATATGAGAGCTCTC 866
Db 3108 LeuGlyAlaSerValGlnProProValAlaLeuProGlyAlaIleGlyThrGlySerVal 3127
QY 867 CCATCAGCTCTCAATGGGCTGCTTTCCTCTGGACCT-----905
Db 3128 ProGlyAla-----GlyGlyTrpGlyProTrpGlyProTrpSerHisCysSerArgSer 3145
QY 906 -----GCTGCAAGGGCCACCCCAATAAATCTG 935
Db 3146 CysGlyGlyGlyLeuArgSerArgThrArgAlaCysAspGlnProProGlnGlyLeu 3165
QY 936 TCTGGGTACTGTGAC-----TGCCTTCTCAGGGGAGACTCTGCAACAGC-----980
Db 3166 GlyAspTyrCysGluGlyProArgAlaGlnGlyGluValCysGlnAlaLeuProCysPro 3185
QY 981 -----TGC-----983
|||
```

Db 3186 ValThrAsnCysThrAlaIleGluGlyAlaGluTyrSerProCysGlyProCysPro 3205
QY 984 ---AGCTGCAACAACTCGCCATCAGCTCGAGCGCTTCAAGCCATAAAGCGGTGCTT 1040
Db 3206 ArgSerCysAspAspLeuValHis-----CysVal 3215
QY 1041 GATAGA---AATCCCTGAAGCTTCCAAACCAAAA----- 1070
Db 3216 TrpArgCysGlnProGlyCysTyrCysProProGlyGlnValLeuSerSerAsnGlyAla 3235
QY 1071 -----ATGGGGAAGCGCTGGAGCTGCTAAACTT-----CGACACAGC 1112
Db 3236 IleCysValGlnProGlyHisCysSerCysLeuAspLeuLeuThrGlyGlnArgHis 3255
QY 1113 AAAGGTGCAACTGTAAGCGC---TCAGGCTGCCTGAAGAACTACTGTGAGTGCTATCAG 1169
Db 3256 ProGlyAlaArgLeuAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGlu 3273
QY 1170 GCCAAATCATGTGTTCTTCCATTTGCCAATGTCATTGCTGCAAAACTATGAAGAAAGT 1229
Db 3274 GlyArgLeuAsnCysThrAspLeu---ProCysProAspCys----- 3286
QY 1230 CCAGAACCAAAATCTGATGAGCACACCCACTACATGGAGCCTGGGAGCTTTGAGAGC 1289
Db 3287 -----GlyGlyGlyGlnSer 3291
QY 1290 AGCCATTATTTGCTCCCGCAGCAAGTCTCAGGACCTCCAAACTCAGAAAAATAGGCAG 1349
Db 3292 LeuHis-ProCysGlyGlnProCysProArgSerCysGlnAsp----- 3305
QY 1350 GCCTTCTCTCTATCTCCTCGGAAGTAGTGGAGGCACATGTGCTGCTGCTGCCGCC--- 1407
Db 3306 -----LeuSerProGlySerValCysGlnProGlySerValGlyCysGlnProth 3322
QY 1407 ----- 1407
Db 3322 rCysGlyCysProLeuGlyGlnLeuSerGlnAspGlyLeuCysValProProAlaHisCys 3342
QY 1408 -AGGCTAGGAAGCAGACAGGAGCAGCTGTTCCCA----- 1443
Db 3342 sArgCysGlnTyrGlnProGlyAlaMetAlaProSerPheValProSerThrCysValAl 3362
QY 1444 -----GCTTGCTGAGCAGC 1457
Db 3362 aclyIleLeuGlnCysGlnGluValProAspCysProAspProGlyValTrpSerSerTr 3382
QY 1458 ATGATCCTGGAGGATTTGGAAAGTGCTGTCGCGAGATTCACATCGAGTTCAAGTCC 1517
Db 3382 pGlyProTrpGluAspCysSerValSerCys-----G1 3393
QY 1518 AAGGGCTGAAATAGTAGTACGCTGCAAGCTGGTAAAGGGGAATGCCTGTGCAAGCCTC 1577
Db 3393 yGlyGlyGlnLeuArgSerArgArgCysAlaArgProProCysProGlyProAlaArg 3413
QY 1578 AGCCCTGGGAATCTGCACCGAGGAGCTGGTCCCGAGGAGCAGAGCGCGCATCA 1637
Db 3413 gGlnSerArgThrCysSerThrGlnValCysArgGluAlaGlyCysProAlaGlyArg-- 3432
QY 1638 TGGCCAGGTCAGCTGAGGCTGAGTGTGATGTCATGTCATGTCGTCGTCGTCGTCGTC 1697
Db 3432 ----- 3432
QY 1698 ATCCTAAAGTCAACGAGCAGCAGCACCCTGGGAGTGACACTGGCCCTCTGTCCTCTG 1757
Db 3433 -----LeuTyrArgGluCysGlnProGlyGlyGly-----CysProPh 3445
QY 1758 G-----GGAGGCCCTCTCGGGAC 1775
Db 3445 eSerCysAlaHisValThrGlnGlnValGlyCysPheSerGluGlyCysGluGluGlyCy 3465
QY 1776 TCCCTGCCCTGCATAAAAAGAGGGTATTTCTACTGTTGTTATGTTGTTGTTCTTCAA 1835
Db 3465 sHisCysPro-----GluGlyThrPheGlnHisArgLeuAlaCys---ValGlnG1 3481

QY 1836 TTGCTTAGTAGTACCTCCATTCAGTTTATGAGCCAGCCTCAAGTTTAGAGAGTAGGC 1895
Db 3481 uCysProCysVal-----LeuThrAlaTrpLeu--LeuGlnGluLeuGly 3495
QY 1896 TCTTCTTTCAGGTGGACTCTGCCCAATCACATACAAAGTCAGGTGCCCATCAGGGGTTTTT 1955
Db 3496 AlaThrIleGlyAsp-----ProGlyGln-----ProLeuGly----- 3506
QY 1956 CCAGCCAGGCTGTGACAGGAGATATGGAGGGGGTTCGGGTTAGAGCTGGGTTGTTT 2015
Db 3507 ProGlyAspGluLeuAspSer-----GlyGlnThrLeuArgThrSerCysGly----- 3522
QY 2016 GGATTTTTTGCCTTTTTTCTTCTGCTATTTCTGCTGAAGTGAGAAAACTTGTCTCTCTG 2075
Db 3523 -----AsnCysSerCysAlaHisGlyLeuSerCysSerLeu 3535
QY 2076 TCCAACTTTTCTCCATAATTACTGCTGACGGTGCCTGTGACCAAGTCACAGTGAAGCT 2135
Db 3536 AspAspCysPheGlu-----AlaAspGlyGlyPheGlyPro 3547
QY 2136 CAGACACAGAGGTGAGGTGGCTTATTATGCCCCACACTTTG----- 2177
Db 3548 TrpSerProTrpGlyProCysSerArgSerCysGlyGlyLeuGlyThrArgThrArgSer 3567
QY 2178 -----TCGTTTTGTGTGAGATAAACCTTCCACACTCCCA 2213
Db 3568 ArgGlnCysValLeuThrMetProThrLeuSerGluLeuPro 3581

RESULT 8
US-10-189-971-6
; Sequence 6, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-6

Alignment Scores:
Pred. No.: 0.0689 Length: 1057
Score: 125.50 Matches: 149
Percent Similarity: 27.56% Conservative: 66
Best Local Similarity: 19.10% Mismatches: 267
Query Match: 3.13% Indels: 299
DB: 9 Gaps: 37

US-09-743-237-1 (1-2241) x US-10-189-971-6 (1-1057)

QY 374 AGAAGCTCTGATTCCTGCTCTACATCCTGAGTGTACAGGGGAAATCATTT 433
Db 41 ArgThrSerProThrProLeuThrProAlaValCysValAlaValSerGlyAsnValGln 60
QY 434 TTGGTGACACTCCGATGACTACTGCCAGGTTCACAGCAGCAGCAAGCAAGCAAGAAAG 493
Db 61 CysLeuAlaArgArgCysValProLeuProCysProGluProValLeuLeuProGlyGlu 80
QY 494 TGTTCAATCAAGAAGACAGGTGGTAGTGTGCCAGCGCGCAGCCCTGAAGACGACGCTTT 553

RESULT 9

```

US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication NO. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20030028907A1el Human Kielin-like Proteins and Polynucleotides
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

```

Db 817 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 837
QY 1731 -----GGATGGACACTG----- 1742
Db 837 nGlyArgGluValAlaSerGlyGluArgTTPThrValAspThrCysThrSerCysSerCy 857
QY 1743 -----GCCTCTCTGTCCTGGGAGGCC 1765
Db 857 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 875
QY 1766 CTCTGGGGACTCCCTGCCCTGCATATAAAAGAGGGTGTATTTCTACTGTGTATGTGT 1825
Db 876 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 899
QY 1826 TGCTTTCAAAATGCTTAGTAGTACTCCTCAATCAAGTTATTATGAGCCAGCCTCAAGTTAG 1885
Db 890 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 907
QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGGACTCTGCCCAATACATACAAAGTCAG 1936
Db 907 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 927
QY 1937 GTGCCATCAGGGTTTTTCAGGCCAGGCTGTGACAGGAGATATGGGAGGGGTCGG 1996
Db 927 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 946
QY 1997 GTTAGAGCTGG 2007
Db 946 lyValAlaTTP 949

RESULT 10

US-10-189-971-24

; Sequence 24, Application US/10189971

; Publication No. US20030028907A1

; GENERAL INFORMATION:

; APPLICANT: Walker, D. Wade

; APPLICANT: Scoville, John

; APPLICANT: Turner, C. Alexander Jr.

; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleot

; TITLE OF INVENTION: Same

; FILE REFERENCE: LEX-0360-USA

; CURRENT APPLICATION NUMBER: US/10/189, 971

; CURRENT FILING DATE: 2002-07-03

; PRIOR APPLICATION NUMBER: US 60/302,949

; PRIOR FILING DATE: 2001-07-03

; PRIOR APPLICATION NUMBER: US 60/315,634

; PRIOR FILING DATE: 2001-08-29

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 24

; LENGTH: 1342

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-24

Alignment Scores:

Pred. No.:	0.0991	Length:	1342
Score:	124.00	Matches:	135
Percent Similarity:	28.32%	Conservative:	59
Best Local Similarity:	19.71%	Mismatches:	227
Query Match:	3.09%	Indels:	265
DB:	9	Gaps:	36

US-09-743-237-1 (1-2241) x US-10-189-971-24 (1-1342)

QY 558 GCCCTCTGGCT-----CAGGAATCTGTTC----- 584
Db 420 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 439
QY 585 -----AAGTTCCTCATCATCCAGGAGCA----- 608
Db 440 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 459

QY 609 -----GAGGAGGCCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTG 656
Db 460 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 479
QY 657 ATTGTGACGTGAAGAGGCGCCAGATGCTCTGCATAGACAACACTGTCGCGGAGGAG 716
Db 480 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 495
QY 717 CTCAAGAGCGCTCATCTGCTTCTCAGTACGATGACCAGAGCAGTTCCTCAGTCAGAG 776
Db 496 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 511
QY 777 CTCCTTAAGCCAATGACAACCTTTAGTGGGAAGACTTCTGCCA----- 818
Db 512 GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys 531
QY 819 -----GTACCAGCG----- 827
Db 532 GluProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln 551
QY 828 -----AAGTTAAATCTCATCACACAGGTGTGATGAGAGCTCTCCATCAGCTGCAAT 881
Db 552 GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp 571
QY 882 GGGGCTGCCTTT----- 899
Db 572 ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys 591
QY 900 GGACCTGCTCTGCAAGGGCCACCCCAAAATAACTCTGTCTGGGTACTGTGACTGCTTCTCC 959
Db 592 AlaProAlaLeuCysProHisProSer-----ProGlyProCysPheCys----- 606
QY 960 AGCGGGGACTTCTGCAACAGCTGCTGCAACAACCTGCGCCATGAGCTCGAGCGCTTC 1019
Db 607 -----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln----- 620
QY 1020 AAAGCCATAAAGCGTGTCTTTGATAGAAATCTCGAAGCTTTCCAAACCAAAATGGGAAA 1079
Db 621 -----AspGlyGluGluPheGluGlyProAlaGlySer 631
QY 1080 -----GGCGTCTGGGAGCTCTAAACTCGA----- 1106
Db 632 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 651
QY 1107 -----CACAGCAAGGGTGCACAACTGTAAGCGC---TCAGGCTGC 1142
Db 652 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 671
QY 1143 CTGAAG----- 1148
Db 672 LeuAlaHisGlyGluGluHisProGluGlySerArgTyrValProProAspSerAlaCys 691
QY 1149 AACTACTGTGCTGCTATGAG-----GCCAAATCATGTCTTCTTCCATT 1193
Db 692 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 711
QY 1194 TGCAAA-----TGCAATTCCTTGCAAAACATATGAAGA 1226
Db 712 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 731
QY 1227 AGTCCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGAGCGCTGGGAC-----TTT 1283
Db 732 GluGlyArgLys-----TyrGluProGlyGluSerPhe 742
QY 1284 GAGAGCAGCATATTGTGTCCTCCAGCCCAAGTTC-----TCAGGA 1322
Db 743 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 760
QY 1323 CCTTCAAACTGAGAAAAATAGCGGCTTCTCTCTGATATCTCTCTGGGAGTAGTGAG 1382
Db 761 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 776
QY 1383 GCCACATGTGCTGCTGCTGCGCCAGGGGTGAGGAAGCAGCAGGAGGACTGTTCCCA 1442

Best Local Similarity: 19.71%				Mismatches: 227			
Query Match: 3.09%				Indels: 265			
DB:				Gaps: 36			
US-09-743-237-1 (1-2241) x US-10-189-971-2 (1-1628)							
QY	558	GCCTCTGGCT-----	CAGGAATCCTGTTGC-----	584			
Db	706	AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr	725				
QY	585	-----	AAGTTCCATCATCCAGGAGCA-----	608			
Db	726	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu	745				
QY	609	-----	GAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGGCCCATGGTG	656			
Db	746	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys	765				
QY	657	ATTGTGTCAGCTGAAGAGGCGCCAGATGCTCTGCATAGACAACTGTGCGCGAGGGAG	716				
Db	766	ProPheProAlaArgGlyAsp-----	781				
QY	717	CTCAAGCGCTCATCTGCTTCCTCAGTACGATCACCAGAGCAGTTTCCCTCAGTCAGAG	776				
Db	782	-----	TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	797			
QY	777	CTCCCTAAGCCAATGACAACTTTAGTGGGAAGACTTCTGCCA-----	818				
Db	798	GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys	817				
QY	819	-----	GTACCAGCG-----	827			
Db	818	GluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln	837				
QY	828	-----	AAGTTAATCTCATCACACAGGTGATAATGGAGCTCTCCCATCAGCTGTCAAT	881			
Db	838	GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp	857				
QY	882	GGGCTGCCTTT-----	CCCTCT	899			
Db	858	ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys	877				
QY	900	GGACTGCTCTGAAGGGCCACCAATAACTCTGCTGGGACTGTGACTGCTCTCC	959				
Db	878	AlaProAlaLeuCysProHisProSer-----	ProGlyProCysPheCys	892			
QY	960	AGCGGGGACTTCTCAACAGCTCAGCTGCACAACTGGCCCATGAGCTCGAGCGCTTC	1019				
Db	893	-----	ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln-----	906			
QY	1020	AAAGCCATAAAGCGGTCTTGTATAGAAATCCTGAAGCTTTCACCAACCAAAATGGGAAA	1079				
Db	907	-----	AspGlyGluGluPheGluGlyProAlaGlySer	917			
QY	1080	-----	GGCGCTCTGGAGCTGCTAACTTCA-----	1106			
Db	918	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro	937				
QY	1107	-----	CACACGAAGGGTGAACCTGTAAGCGC-----TCAGGCTGC	1142			
Db	938	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	957				
QY	1143	CTGAAG-----		1148			
Db	958	LeuAlaHisGlyGluGluHisProGluGlySerArgTyrValProProAspSerAlaCys	977				
QY	1149	AACTACTGTGACTATGAG-----	GCAAAATCATGTGTTCTTCCATT	1193			
Db	978	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer	997				
QY	1194	TGCAAA-----	TGCATTGCTGCACAAAACATCAAGAA	1226			
Db	998	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis	1017				

RESULT 14

US-10-025-380-1116

; Sequence 1116, Application US/10025380

; Publication No. US20020182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.

; APPLICANT: Wang, Tongcong

; APPLICANT: Jiang, Yudi

; APPLICANT: Smith, Carole L.

; APPLICANT: King, Gordon E.

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan D.

; APPLICANT: Skeiky, Yasir A. W.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Vedwick, Thomas S.

; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.471C14

; CURRENT APPLICATION NUMBER: US/10/025.380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1116

; LENGTH: 5405

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1116

Alignment Scores:

Pred. No.:	0.368	Length:	5405
Score:	119.50	Matches:	122
Percent Similarity:	31.83%	Conservative:	55
Best Local Similarity:	21.94%	Mismatches:	162
Query Match:	2.95%	Indels:	218
Db:	9	Gaps:	34

US-09-743-237-1 (1-2241) x US-10-025-380-1116 (1-5405)

Qy	1479	TTCCAACTCTCCAGGATCATCTGCT-----CAGCCAAAGCTTGGGACAGT-----	1432
Db	1401	PhGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu	1420
Qy	1431	-----GCTCCTGCTGCTTCTCCTCACCCTGGCCACGAGGC-----	1396
Db	1421	GluValValProAsnSerProCysLeuProProThrProCysProProGlySerGluAsp	1440
Qy	1395	-----AGGCACATGGGCTCCACTACTTCCAGGAGATACAGGAGAGGCTGCC	1345
Db	1441	CysIleProSerHisLysCysProGluLeuGluLysLysTyrglnLysGlu-----	1458
Qy	1344	TATTTTCTCAGTTTGGAGTCTGAGACTTGGCTGGGGACAAATAATGGCTGCTCT	1285
Db	1459	-----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis	1475
Qy	1284	CAA-----AGTCCCAAGCTCCCA-----TGTAAGTGGG-----	1258
Db	1476	LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly	1495
Qy	1257	-----GTGTGCTCATCAGCATTTTTCGTTCTGGACTTCTTCCATAGTTT	1213
Db	1496	GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTyrr	1508
Qy	1212	TCCAAGCAATGC-----ATTTCAAATGGAAGACATGATTGG	1171
Db	1509	valSerAlaCysGlnAlaAlaGlyHisValGluProTrpTrpThrGluThrPheCys	1528
Qy	1170	CCTATAGACTCAGAGTAGTTCTTCAGGAGCCTTGAGCGCTTACAGTTGACCCCTTGC	1111

Db	1529	ProMetGluCysProProAsnSerHisTyrgluLeuCysAlaAspThrCysSerLeuGly	1548
Qy	1110	TGTGTGCAAGTT-----TAGCAGCTCCAGACGGGCTT-----	1078
Db	1549	CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys	1568
Qy	1077	---TCCCATTTTGTGTGGAAAGCTTCAGGATTCTATCAAGACACGCTTTATGGCTT	1021
Db	1569	AspSerGlyPheLeuTyrr-----	1574
Qy	1020	TGAAGCGCTCGAGCTCATGGCGCAGTTGTGTC-----AGCTGACAGCTGTTGAGA	970
Db	1575	-----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyrrHis	1589
Qy	969	AGTCCCGCTGGAGAGCAGTCACAGTACCCAGACAGAGTTATTTGGGTGGCCCTTGA	910
Db	1590	Asn-----	1590
Qy	909	GAGCAGTCCAGAGGAAAGGAGCCCATTCACAGCTGATGGGAGAGCTCCATTATCAA	850
Db	1591	-----GlyValTyrrTyrrGlu	1595
Qy	849	CCTGTGTGATGAGATTTAACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	790
Db	1596	Pro-----GluGlnThrValLeulle-----	1602
Qy	789	TTGGCTTAGGAGCTGCTGACTGAGGAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT	730
Db	1603	-----AspAsnCysArgGlnGlnCysThrCysHisAlaGly	1614
Qy	729	GGAGCGCTTGGAGTCCCTCGCGCCACAGTCTG-----CTATGACAGAGCATCT	682
Db	1615	LysGlyMet-----ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer	1633
Qy	681	GGGCGCTCCTTTTCA---GCTGACAAATCACCATGGGCTGGAGTCTTCTTCCGAGGC	625
Db	1634	GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly	1647
Qy	624	---AGCTGGAGCGCTCCT-----CTGCTCTCTGGGATGATGGAACTTGCACACAG	577
Db	1648	ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlnGlyValCys-----	1665
Qy	576	ATTCTGTAGCCAGAGGGCGCTGAAAGCTGCTCTCAGGGCTGCCCGCTG-----	526
Db	1666	LeuProAsnTyrrGluAlaThr-----CysTrpLeuTrpGlyAsp	1678
Qy	525	GCACACTACCACTGCTTC-----TTTGATTTC---AACACTTTTCTTGTCT	481
Db	1679	ProHisTyrrHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyrrVa	1698
Qy	480	GCTTGTGCTGGGAGCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	422
Db	1698	IleuAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr	1714
Qy	421	CTGTACACCTACCACTCAGGATGTAGAGCAGAGAAATCAGAGTCTCTCTAGTTATAT	362
Db	1714	OpheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyrrVa	1732
Qy	361	AGCAATTCAGGAATTCATATATAGCTTAGCAATATG---GTTGAGAGACAGAAATTCAT	305
Db	1732	IargValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluIleG	1752
Qy	304	CAAACTCTACATAGT-----TTAAGAACTGTCCCTTATCTTACAGAGGGGAGAT	251
Db	1752	LyLysValArgValAsnGlyValLeuThrAlaLeuProValSerValAlaAspGlyArgI	1772
Qy	250	CATCAACTGAAAAATCCTCAATATGGAAGTCTTGGTGTGAAGAAAAAATATCCTCC	191
Db	1772	IeSerVal-----	1774
Qy	190	AGCCAAACCCAACTACCAGGCTGAAGACTATCTGCAGAACACGCAAAAGGCCTGGTT	131

Db 1775 -----ThrGlnGlyAlaSerLysAlaLeu----- 1782
QY 130 CTATCCCTAGCCCTGCTGGGGAGGTGGT-----GGAGTCA 92
Db 1783 --LeuValAlaAspPheGlyLeuGlnValSerTyrAspTrpAsnTrpArgValAspValt 1802
QY 91 CAAGTCCCAAGCGCCAGGAACACACACTGTGTATCCCTTGTGGT 48
Db 1802 hrLeuProSerSerTyrHisGlyAlaValCysGlyLeuCysGly 1816
RESULT 15
US-09-922-217-1116
; Sequence 1116, Application US/09922217
; Patent No. US20026076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1116
; LENGTH: 5405
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1116
Alignment Scores:
Pred. No.: 0.368 Length: 5405
Score: 119.50 Matches: 122
Percent Similarity: 31.83% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
DB: 10 Gaps: 34
US-09-743-237-1 (1-2241) x US-09-922-217-1116 (1-5405)
QY 1479 TTCAAACCTCTCCAGGATCATCTGCT-----CAGCCAAAGCTGGGGAAACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCCTGCTCTGCTTCTCCACCTCGCCGCGCAGCAGC----- 1396
Db 1421 GluValValProAspSerProCysLeuProThrProCysProGlySerGluAsp 1440
QY 1395 -----AGGCATCTGGCTCCACTACTCCAGGAGATACAGGAAGCCCTGCC 1345
Db 1441 CysIleProSerHisLysCysProProGluLeuGlyLysTyrGlnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGGCTCTCAGAACTTGGCTGGGCAATAATGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCCAGGCTCCA-----TGACTGGG----- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495
QY 1257 -----GTCGCTCATCAGCATTTTTCGCTTCGACATTTTCATAGTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTyr 1508

QY 1212 TGCAAGCAATGC-----ATTGCAATGGAAGAACACATGATTTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValCluProTrpArgThrGluThrPheCys 1528
QY 1170 CCTCATAGCACTCACAGTAGTCTTCTAGGAGCGCTGAGCGCTTACAGTTGCACCCCTTGGC 1111
Db 1529 ProMetGluCysProProAsnSerHisTyrGluLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TCTGTCGAAGTT-----TAGCAGCTCCAGAGCGCCCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCATTTTGGTTGGAAAGCTTCAGGATTTCTATCAAGACACGCTTTATGGCTT 1021
Db 1569 AspSerGlyPheLeuTyr-----1574
QY 1020 TGAAGCGCTCGAGCTCATGCGCAGGTGTTC-----AGCTGCAGCTGTTGCAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyrHis 1589
QY 969 AGTCCCGCTGGAGAAGCAGTCACAGTACCAGACAGAGTTATTTGGGTGGCCCTTGCA 910
Db 1590 Asn-----1590
QY 909 GAGCAGGTCAGAGGAAAGGAGCCCATTTGACAGTGTGGAGAGCTCCATTATCAA 850
Db 1591 -----GlyValTyrTyrGlu 1595
QY 849 CCTGTGTGATGAGATTTAACTTCGCTGGTACTGGCAGAGCTCTCCACATAAGTTGTCA 790
Db 1596 Pro-----GluGlnThrValLeu-----1602
QY 789 TTGCCTTAGGAGCTCTGACTGAGGGAACCTGCTCTGTCATCTACTGAGGAAGCAGAT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTTGAGCTCCCTCGCCACAGATGT-----CTATGCAGCATCT 682
Db 1615 LysGlyMet---ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTTTCA---GCTGACAAATCACATGGGCGTGGAGTCTTTCTCCGAGGCG 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGGCTCCT-----CTGCCTCTGGGATGATGGAACTTGCACAGG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlnGlyValCys----- 1665
QY 576 ATTCTTGAGCCAGAGGGGCTGGAAAGCTGCGTCTTCAGGGCTGCCCTG----- 526
Db 1666 LeuProAsnTyrGluAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACATACACCTGCTTC-----TTTGATTTC---AACACTTTTCTTCTGCTT 481
Db 1679 ProHisTyrHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyrVa 1698
QY 480 GCTTGTGCTTGGAACTTGGCAGTGTATCATCGGAGTCTACCAAAATG-ATTTCGCC 422
Db 1698 lLeuAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTTAGCAGCTAGGATGTAGAGGAGGAGATCAGAGTTCTCTCAAGTTATAT 362
Db 1714 oPheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyrVa 1732
QY 361 AGCAATTTCAAGAAATGCTATATAGCTTAGCAATATG---GTTGAGAACAAGATTCAT 305
Db 1732 lArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspLulleg 1752
QY 304 CAAACTCTACATAGT-----TTAAAGAACTGCTCCCTTATCTTACAGAGGGGGAGAT 251
Db 1752 lLysValArgValAsnGlyValLeuThrAlaLeuProValSerValAlaAspGlyArgI 1772

```
QY 250 CATCAACTGAAAAATCCTCAATAATGGAAGTCTTTGGTGTGAAGAAAAAAATCTCC 191
Db 1772 leSerVal-----1774
QY 190 AGCCAAACCCAACTACCAAGGCTGAAGACTATCTCAGAACACGCAAAAGGCCCTGGGTT 131
Db 1775 -----ThrGlnGlyAlaSerLysAlaLeu----1782
QY 130 CTATCCCTAGCCCTGCTGGGGGAGGTGGT-----GGAGTCA 92
Db 1783 --LeuValAlaAspPheGlyLeuGlnValSerTyrAspTrpAsnTrpArgValAspValT 1802
QY 91 CAAGTCCCAACAGGCCAGGAACACACACTGTGTATCCCTTGTGGT 48
Db 1802 hrLeuProSerSertyrHisGlyAlaValCysGlyLeuCysGly 1816
```

Search completed: April 21, 2003, 11:48:24
Job time : 124.359 secs


```
QY 525 -----CCAGGCGCAGCCCTGAAGCGCAG----- 549
Db 42 oAlaGlnGlnAsnArgProGlyThrSerSer-MetIleGlnGlnHisAsnArgSerMetG 62
QY 550 --CTTTCCAGCGCCCTCTGGCTCAGGAATCCTGTGTGCAAGTTCCTCCATCAT- 597
Db 62 luValasngInGlyLeuValysAspCluProIleasphThrSerHisValtyrV 82
QY 598 -----CCAGAGGCGCAGGAGGCTCCAGCTGCCCTCGGAAAGAAAGACTCCAGCCC 649
Db 82 alProProProArgProValGlnArgLysPro-----GlyProSerThrPro--- 97
QY 650 CATGGTGATTGTTCAGCTGAAGAGGCGCCAGATGCTCTGCATAGACAACTGTGGCGC 709
Db 97 ----- 97
QY 710 GAGGAGCTCAAGCGCTCATCTGCTTCTCAGTAGCAGTACCAGAGCAGATTTCCTCA 769
Db 98 --GlySerSerGlnTyrThrValArgAsnLeuSerAsnLeuSerGlySerProSerMet 117
QY 770 G-----TCAGAGCTCCCTAAG-----CCAATGACAACCTTGTAGTGGAA 807
Db 117 yrAspArgGlnProAlaSerLeuProArgThrValGlnPrometGlyLeuGluMetGlyA 137
QY 808 GACTTCTGCCAGTACCAGCGAAGTTAAATCTCATCAGAGTTGATATGGAGCTCTCC 867
Db 137 snSerGluGlnArgLysValTyrIleAspMetLysAspHisValSerHisIleArgLeu 157
QY 868 CATCAGCTGTCAATGGGCTGCTTTCCTCTGAGCTGCTCTGCAAGGCGCCACCCAAA 927
Db 157 ysthrLysLysValPheAla-----ProGlyGlnArgLysProCysAsnC 173
QY 928 TAAC-----CTGTCTGGTACTGTGACTTCTCCAGCGGGAGCTTCTGCA 975
Db 173 ysthrLysSerGlnCysLeuLysLeuTyrCysAspCysPheAlaAsnGlyGluPheCysA 193
QY 976 ACAGCTGCAGCTGC-----AACAACTGCGCCATGAGCTCGAGCGCTTCAAAG 1023
Db 193 rgAspCysnGlyLysLysAspCysHisAsnIleGluTyrAspSerGlnArgSerLysA 213
QY 1024 CCATAAGCGCTGCTGTATAGAAATCTCGAAGCTTTCACCAACAAATATGGGAAAGGCC 1083
Db 213 laileArgGlnSerLeuGluArgAsnProAsnAlaPheLysProLysIleGlyIleAlaA 233
QY 1084 GTCTGGAGCTGCTAACTTCGA-----CACAGAAAGGTCGAACCTGTAAGCGCTCAG 1137
Db 233 rgGlyGlyIleThrAspIleGluArgLeuHisGlnLysGlyCysHisCysLysSerG 253
QY 1138 GCTGCTGAAGAACTACTGTGAGTGTATGAGCCAAATCATGTGTTCTTCTCCATTGCA 1197
Db 253 lyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysValProCysThrAspArgCysL 273
QY 1198 AATGCATTGCTTGCAAAAC-----TATGAAGAAAGT----- 1229
Db 273 yscysLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSerGlyG 293
QY 1230 -----CCAGACGAAATATGCTGATGAGC-----ACAC 1257
Db 293 lyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAsnAlaSerThrAlaThrp 313
QY 1258 CCCACTACATGAGCTGGGACTTTGAGAGCAGCCATTAT-----TTGTCCCCAG 1308
Db 313 roAsp---SerGlyProGlySerValThrAspGluHisGlyAspAspTyrGluAspM 332
QY 1309 CCAAGTCTCAGGACTCCAAACTGAGAAAAATAGCAGCGCTTCTCTCTATCTCTCT 1368
Db 332 etLeuLeuSerHisLysProLysValGluMetAspProArgArgPhe-----Prot 349
QY 1369 GG-----GAAGTAGTGGAGGCACATGCTGCTGCTGCGCCAGGGTG 1413
Db 349 rpyTyrMetThrAspGluValValGluAlaAlaThrMetCysMetValAlaGlnAlaG 369
QY 1414 AGGAAGCA-----GAGCAGGAGCACTGTTCCCAAGCTGGCTGAGC 1455
```

```
Db 369 luGluAlaLeuAsnTyrGluLysValGlnThrGluAspGluLysLeuIleAsnMetGluL 389
QY 1456 AGATGATCTGGAGGAGTTTGGAAAGTGCTGTCGCAGAGATTCTC 1499
Db 389 ysLeuValLeuArgGluPheGlyArgCysLeuGluGlnMetIle 403
RESULT 2
B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventre
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: B84420; MUID:20083487; PMID:10617197
A:Accession: B84585
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
A:Cross-references: GB:AE002093; NID:g4580462; PIDN:AAD24386.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g20110
A:Map position: 2
```

```
Alignment Scores:
Pred. No.: 2,89e-22 Length: 571
Score: 371.00 Matches: 92
Percent Similarity: 45.85% Conservative: 46
Best Local Similarity: 30.56% Mismatches: 101
Query Match: 9.26% Indels: 62
DB: 2 Gaps: 11
```

US-09-743-237-1 (1-2241) x B84585 (1-571)

```
QY 603 GAGGACAGGAGGAGCCCTCCAGCTGCCCTCGGAAGAAAGACTCCAGGCCCATGGTATTGT 662
Db 3 GluGlyGluGluGlyAspLysPheProProLysThrAspGlu-----ValThrGln 19
QY 663 CAGCTGAAAGGAGGCGCCAGATGCTGCTCATAGACAACGTGTGGCGGAGGAGCTCAA 722
Db 20 GluSerMetLysSerAlaArgGlnLeuAspPheThrGly---GlySerSerValGlu 38
QY 723 GCGCTCCATCTGCTCTCCTCAGTACGATCACCAGAGCAGTTTCCCTCAGTCAGAGCTCCCT 782
Db 39 HisSerHis-----SerAsnGlnAlaSerSerMetAlaAlaSerIlePro 54
QY 783 AAGCCAATGACAACCTTACTGGGAAGACTTCTG----- 815
Db 55 SerProIleValThrValThrArgProIleIleThrSerGlnAlaProThrValAla 74
QY 816 -----CCAGTACCAGCAGAGTTAAATCTCATCATCACAGGTTTCAT---AATGA 860
Db 75 ThrProIleProProProProGlnSerGlnGlyIleIleLeuHisValProIleArgHis 94
QY 861 GCTCTCCCATCAGCTGCTCAATGGGCTGCTTTCCTCTGAGCTGCTGCTGCAAGGCCA 920
Db 95 ProArgProGluSerProAsnSerMetProArgProAlaGlyGluThrArgAspGlyThr 114
QY 921 CCCAAA-----ATAACTCTCTCTGGGTACTGTGAC 950
Db 115 ProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLysLeuTyrCysGlu 134
QY 951 TCGTTCTCCAGCGGGAGCTTCTGCAACAGCTGCAGCTGC-----AACACCTG 998
Db 135 CysPheAlaSerGlyThrTyrCysAspGlyCysAsnCysValAsnCysPheAsnVal 154
QY 999 CCGCATCTCGAGCGCTTCAAAGCCATAAAGCGGTGTCTGTATAGAAATCTGAAGCT 1058
Db 155 GluAsnGluProAlaArgGlnAlaValGluSerThrLeuGluArgAsnProAsnAla 174
```

Qy	1059	TTCCACCAAAAATGGGAAAGCGCTGTGGGAGCTGCTATAA-----	1100
Db	175	PheArgProLysIleAlaSerProHisGlyArgAspAsnArGluGluValGly	194
Qy	1101	-CCTTCGCACAGCAAAGGTCGAACCTTAAGCGCTCAGCGTGCCTG	1145
Db	195	AspValMetLeuAlaArgHisasnlysGlyCysHIScysLysLysSerGlyCysLeu	214
Qy	1146	AAGAACTACTTGAGTGCATTATGAGGCCAAATCATGTCTTCCTCATTTGCAAAATGCATT	1205
Db	215	LysLysTyrcysglucypsheglnlaasnilleLeuCysSerGluAsnCysLysCysLeu	234
Qy	1206	GCTTGC AAAA ACTATG AACA AGAG TC CAGAAC GAAAA ATG CTGATG AGCAC ACC CC AC---	1262
Db	235	AspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHisGlyGluHisSer	254
Qy	1263	---TACATGCAG- : : -	1292
Db	255	HisAsnLeuAlaTyrlenglnHisalaSnAlaAlalleThrGlyAlaIleGlySerSer	274
Qy	1293	CATTATTGTC CCC AGCC AA GTTCTCAGGACCTCCAAAAC TGA GAAAA ATA TAGCAGGCC	1352
Db	275	GlyPheAlaSer-----AlaProProLysArgArgLysGlyGlnGluIle	290
Qy	1353	TTC 1355 	
Db	291	Phe 291	
RESULT 3			
T08955			
hypothetical protein F19B15_30 - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999			
C:Accession: T08955			
R:Bavan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft,			
submitted to The Protein Sequence Database, May 1999			
A:Reference number: Z16519			
A:Accession: T08955			
A:Molecule type: DNA			
A:Residues: 1-603 <BEV>			
A:Xrefs-references: EMBL:AL078470; GenBank:U000062; ATSP:F19B15.30			
A:Crossmatchal source: cultivar Columbia; BAC clone F19B15			
C:Genetics:			
A:Gene: ATSP:F19B15.30			
A:Map position: 4			
A:Introns: 107/2; 144/2; 203/3; 321/3; 356/2; 388/1; 522/1			
Alignment Scores:			
Pred. No.: 7,49e-22 Length: 603			
Score: 366.00 Matches: 99			
Percent Similarity: 47.56% Conservative: 47			
Best Local Similarity: 32.25% Mismatches: 97			
Query Match: 9.13% Indels: 64			
DB: 2 Gaps: 15			
US-09-743-237-1 (1-2241) x T08955 (1-603)			
Qy	585	ANGTTC CATCCAGGAGGAGGAGGCGCTCCAGCTGCCCTCGGAAGAAGACTCC	644
Db	9	GlutProProLysLysaspGlyValGlu----GluGlyPheProThrLysLysProAla	27
Qy	645	AGCCCCATGGTAGTTGTTCAGCTGAAGGAGGCGCCCAGATGCTCTGCATAGACAACTGT	704
Db	28	ArgGlnLeu-----AspPheThrGlyGlySerAspGluHisSerLeuSerLysPro	44
Qy	705	GGCGCGGAGGAGCTCAAGCGCTCCCATCTGCTTCCTCAGTACGATGACCAGCAGATTTC	764
Db	45	AlaAlaProthrValValAlaThrSerValLysProIlelle-----SerSerSerVal	62
Qy	765	CCTCAGTCAGAGCTCC TAAGCCAATGACA CT TTAGTGGGAAGACTTCTGCCAGTACCA	824
Db	63	ProSerThrfileargPro---GlyWetThrIleAlaIleGlyGlnValThrGlnValArg	81

QY	825	GCGAAGTAAATCTCATCACACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCTCAATGGG	884
Db	82	ProThrLeuProMetaIaThrThrMetSerAsnProProSerGlnIleValAsn	101
QY	885	GCT-----GCCTTCCTCTGGACCTGTCTGCAAGGCCA---CCCAAATA	929
Db	102	AlaProIleArgHisProIleProGluSerProLysAlaArgGlyProArgProAsnVal	121
QY	930	-----ACT	932
Db	122	GluGlyArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCys	141
QY	933	CTGTCTGGGTACTGTGACTGCTTCTCCAGCGGGAGCTTCTGCAACAGCTGCGAGCTGC---	989
Db	142	LeuLysLeuTyrcysGluCysPheAlaSerGlyThrTyrcysAspGlyCysAsnCysVal	161
QY	990	-----AACAACTGCGCCATGAGCTCGAGCGGTTCGAAGCCATTAAGCGGTGCTT	1040
Db	162	AsnCysPheAsnAsnValAspAsnGluProAlaArgArgGluAlaValGluAlaThrLeu	181
QY	1041	GATGAATCCTGAAGCTTTCACAACCACAAAATGGGGAAA-----GGCGCT---	1085
Db	182	GluArgAsnProPheAlaPheArgProLysIleAlaSerSerProHisGlyGlyArgAsp	201
QY	1086	-----CTGGGAGCTGTAACATT-----CGACACAGCAAAAGGTGCAACTGT	1127
Db	202	LysArgGluAspIleGlyGluValValLeuLeuGlyLysHisAsnLysGlyCysHisCys	221
QY	1128	AAGCGCTCAGCGTCCCTGAAGAAGCTACTGTGAGTGCTATGAGGCCAAATCATGTGTTCT	1187
Db	222	LysLysSerGlyCysLeuLysLysTyrcysGluCysPheGlnAlaAsnIleLeuCysSer	241
QY	1188	TCCATTTCGAANTGCATTGCTCCAAAACTATGAAGAAGTCCAGAACGAAATGCTG	1247
Db	242	GluAsnCysLysCysLeuAspCysLysAsnPheGluGlySerGluGluArgGlnAlaLeu	261
QY	1248	ATGAGCACACCCCAC-----TACATGGAG-----	1271
Db	262	PheHisGlyGluHisSerAsnHisMetAlaTyrrLeuGlnGlnAlaAlaAsnAlaAlaIle	281
QY	1272	CCTGGGAGCTTTGAGAGCAGCCCATTTATTTG----TCCCCAGCCCAAGTTCTCAGGACCTCCA	1328
Db	282	ThrGlyAlaValGlySerSerGlyPheAlaProSerProAla-----Pro	296
QY	1329	AAACTGAGAAAAATAGGCAG	1349
Db	297	LysArgArgLysGlyGlnGlu	303
RESULT 4			
F71410			
hypothetical protein - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
A:Variety: Columbia			
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998			
R:Accession: F71410			
C:Authors: M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Berck			
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.			
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rec			
C:Chalwatzis, N.			
A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Ara			
A:Reference number: A71400; UID:98121113; PMID:9461215			
A:Accession: F71410			
A>Status: preliminary; nucleic acid sequence not shown; translation not sho			
A:Molecule type: DNA			
A:Residues: 1-658 <BEV>			
A:Cross-references: GB:297337; NID:92244829; PID:e326824; PID:g2244834			
C:Genetics:			
A:Map position: 4COP9-4G3845			

Alignment Scores:

```
Pred. No.: 3.63e-12 Length: 658
Score: 248.50 Matches: 107
Percent Similarity: 39.95% Conservative: 50
Best Local Similarity: 27.23% Mismatches: 160
Query Match: 6.20% Indels: 76
DB: 2 Gaps: 18

US-09-743-237-1 (1-2241) x F71410 (1-658)
QY 441 GACTCCGATGACTACTGCGCA---GGTTCCCAAGCGCAGCAAGCAAGCAAGAAAGTGT 497
Db 186 AspaSerGluLeuLeuIlePheAAspSerProAspAlaSerGluAlaPheArgCysPhe 205
QY 498 GAAATCAAGAAGCAGGTGGTGTAGTGTCCAGGCGGCAGCCCTGAAGACGAGCTTCCAG 557
Db 206 MetMetGlnArgAlaSerAsnSer-----GluAlaArgPheArg 218
QY 558 GCCCCTCTGGCTCAGGAATCCTGTTGCAAGTTCCTATCCAGGAGGAGGAGGCC 617
Db 219 AsnGlyValGluLysGlnThr---MetGlnHisAspSerAsnLysGluProGluSerAla 237
QY 618 TCCAGCTGCCCTCGAAGAAGACTCCAGCCCATGGTGTATGTCAG----- 665
Db 238 AsnAlaIleProTyrGluValAsnSerGly-----ValIleSerGlnAlaValSerLeu 255
QY 666 CTGAAGGAGGCGCCAGATGCTCTGCATAGAC----- 698
Db 256 LeuHisArgGlyIleArgArgArgCysLeuAspPheGluMetProGlyAsnLysGlnThr 275
QY 699 -----AACTGGCGGAGGAGGAGCTCAAGC----- 725
Db 276 SerSerGluAsnAsnThrAlaAlaCysGluSerSerArgCysValValProSerIle 295
QY 726 ---CTCCATCTG-----CTTCCTCAGTACCATCACCAGCAGCTTCCCTCAGTCA 773
Db 296 GlyLeuHisLeuAsnAlaIleLeuMetSerSerLysAspCysLysThrAsnValThrGln 315
QY 774 GAGCTCCTAAGCCAAATGACAACCTTTAGTGGAGAGACTTCTGCCAGTACCAGCAAGTTA 833
Db 316 AspTyrSerCysSerAlaAsnIleGlnValGlyLeuGlnArgSerIleSerThrLeuGln 335
QY 834 AATCTCATACACAGGTGTATATGAGAGCTCTCCCATCAGCTGTCAATGGGCTGCTTT 893
Db 336 AspSerLeuAspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspVal----- 353
QY 894 CCCTCTGAGCTGCTGTGCAA-----GGGCCACCCAAATAAATCTCTGCT 938
Db 354 ProValGluProAlaLeuGlnGluLeuAsnLeuSerSerProLysLys-----Lys 370
QY 939 GGTACTGTGACTGCTTCTCCAGGGGACTTCTGC---AACAGCTGCAGCTGC----- 989
Db 371 SerTyrCysGluCysPheAlaAlaGlyValTyrCysIleGluProCysSerCysIleAsp 390
QY 990 -----AACAACTGCGCCATGACTCGAGGCTTCAAGCCATAAAGCGGTCTTGTAT 1043
Db 391 CysPheAsnLysProIleHisGluAspValValLeuAlaThrArgLysGlnIleGluSer 410
QY 1044 AGAATCCTGAAGCTTTCCACCAAAATGGGAAGAGCCCTCTGGAGCT----- 1094
Db 411 ArgAsnProLeuAlaPheAlaProLysValIleArgAsnSerAspSerValGlnGluThr 430
QY 1095 -----GCTAAACTTCGACACAGCAAGGGTGCAACTGTAAG 1130
Db 431 GlyAspAlaSerLysThrProAlaSerAlaArgHisLysArgGlyCysAsnCysLys 450
QY 1131 CGCTCAGGCTGCCTGAGACACTACTGTAGTGTATGAGGCCAAATATCATGTGTTCTCC 1190
Db 451 LysSerAsnCysLeuLysLysTyrCysGluCysTyrGlnGlyValGlyCysSerIle 470
QY 1191 ATTTGCAATGCATGCTTCCAAAACTATCAGAAAGTCCAGAACCAAAATGCTCATG 1250
Db 471 AsnCysArgCysGluGlyCysLysAsnAlaPheGlyArgLysAspGlySerSerIleAsp 490
```

```
QY 1251 AGCACACCCCATACATGGAGCCTGGGGACTTTGAGAGCAGCCATTATTTGCCAGCC 1310
Db 491 MetGluAlaGluGlnGluGlu-----GluAsnGluThrSerGluLysSerArgThrAla 508
QY 1311 AAGTTCTCAGGACCTCCAAAACACTGAGAAAAATAGGACGAGCCTTCT----- 1356
Db 509 Lys-SerGln-----GlnAsnThrGluValLeuMetArgLysAspMetSerSerAlaLe 526
QY 1357 CCTGTATCTCTCTGGGNAAGTAGTGGAGGCCACATGTCCTGCTGGCCAGGGTGCAG 1415
Db 526 uproThrThrProThrProIleTyrArgProGluLeuValGlnLeu---ProPheSerSe 545
QY 1416 GAAGCAGCAGCAGGACGACTGTTCCCAAGCTTGGCTG 1452
Db 545 rSerLysAsnArgMetProProGlnSerLeuLeu 557

RESULT 5
MEGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motif
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1737 <NA>
A:Cross-references: EMBL:AB011541; NID:G3449307; PIDN:BAA32469.1; PID:G3449308
A:Experimental source: brain; clone HG1392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Alignment Scores:
Pred. No.: 0.00109 Length: 1737
Score: 146.00 Matches: 120
Percent Similarity: 30.86% Conservative: 38
Best Local Similarity: 23.44% Mismatches: 191
Query Match: 3.64% Indels: 163
DB: 2 Gaps: 30

US-09-743-237-1 (1-2241) x T00209 (1-1737)
QY 453 CTACTGCCAGGTTCCTCCAGGACGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 509
Db 609 LeuLeuAlaProSerGlnGlyAlaLysArgAspArgMetArgAsnValArgGlySerSer 628
QY 510 GCAGGT---GGTAGTGTGCCAGCGCGCAGCCCTGAAGACGACGCTTTCACAG----- 557
Db 629 ArgGlyLeuGlyGlnValProGlyGluGlnProGlySerTrpGlyPheArgGluValArg 648
QY 558 -----GCCCTCTGGCTCAGGAATCTGTGTCAGAGTTCCTCATATCC 599
Db 649 LysLysMetAlaLeuTrpAlaAlaLeuAla-----GlyThrGly 661
QY 600 CAGGAGCAGCAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCC----- 650
Db 662 GlyPheLeuGluLysLeuSer-----ProHisLeuLysGluProArgProArgLeuPhe 679
QY 651 -----ATGGTGTATTTGTCAGCTGAAAGGAGGCGCCAG 683
Db 680 HisAlaSerAlaLeuLeuGlyAspThrMetValValLeu-----GlyGlyArgSer 696
QY 684 -----ATGCTCTGCATAGACAACTGTGCGCGAGGAG 716
Db 697 AspProAspGluPheSerSerAspValLeuLeuTyrGlnValAsnCysAsnAla----- 714
QY 717 CTCAAAGCGCTCATCTGCTTCCTCAGTACGATGACCAAGCAGCAGTTCCTCAGTCAGAG 776
```

```
Db 715 -----TrpLeuLeuProAspLeuThrArgSerAlaSer----- 725
QY 777 CTCCTTAAGCAACATTTAGTGGGAAGACTTCTCCAGTACCACGGAAGTTAAAT 836
Db 726 ValGlyProMetGluSerValAlaHisAlaValAlaValGlySerArgLeu 745
QY 837 CTCATCACAGGTTGATTAATAGAGCTCTCCCATCAGCTGTCATAGGGGCTCCCTTTCC 896
Db 746 TyrIleSerGlyGlyPheGlyValAlaLeuGlyArgLeuAlaLeuThrLeuPro 765
QY 897 TCTGCACCTGCT-----CTGCAAGGGCCACCAAAATTAACCTCTGCTGGGTACTGTGAC 950
Db 766 ProAspProCysArgLeuLeuSerSerProGluAlaCysAsnGlnSerGlyAlaCysThr 785
QY 951 ----TGCTTCTCC-----AGCGGGGAC-----TTCTGCAACAGCTGC 983
Db 786 TrpCysHisGlyAlaCysLeuSerGlyAspGlnAlaHisArgLeuGlyCysGlySer 805
QY 984 AGCTGCAACACCTG-----CGCCATGAGCTCGAGCGCTTCAAGGCCATAAGGCG 1034
Db 806 ProCysSerProMetProArgSerProGluGluCysArgArgLeuArgThrCysSerGlu 825
QY 1035 TGTCTGTATAGAAATCTCGAAGCTTTCCACCAAAATGGGGAAGCGCTCTGGAGCT 1094
Db 826 CysLeuAlaArgHisProArgThrLeuGlnProGlyAspGlyGlu----- 840
QY 1095 GCTAAACTTCGACACAGCAAGGTCG---AACTGTAAGCGCTCAGGCTGCCTG----- 1145
Db 841 AlaSerThrProArgCysLysTrpCysThrAsnCysProGluGlyAlaCysIleGlyArg 860
QY 1146 -----AAGAACTACTGT----- 1157
Db 861 AsnGlySerCysThrSerGluAsnAspCysArgIleAsnGlnArgGluValPheTrpAla 880
QY 1158 ----GAGTGTATAGGCCAAATC-----ATGTGTCTTCCATTTCG 1196
Db 881 GlyAsnCysSerGluAlaAlaCysGlyAlaAlaAspCysGluGlnCysThrArgGluGly 900
QY 1197 AAATGCTATGCTTGCAAAACTATGAAGAACTCCAGAA---CGAAATGCTGATGAC 1253
Db 901 LysCysMetTrpArgGlnPheLysArgThrGlyGluThrArgArgIleLeuSerVal 920
QY 1254 ACACCCCACTAC-----ATGAGCGCTGGGGAC 1280
Db 921 GlnProThrTyrAspTrpThrCysPheSerHisSerLeuLeuAsnValSerProMetPro 940
QY 1281 TTGTAGACAGCCATATTTGTCCCGCAGCAAGTCTCAGGACCTCCAAAATCGAGAAA 1340
Db 941 ValGluSerSer-----ProLeuProCysProThrProCysHisLeuLeuPro 957
QY 1341 AATAGCAGGCGCTTCTCCTGTATCTCCTGGGAAGTAGTGGAGGCCACATGTCCTGCCGTG 1400
Db 958 Asn-----CysThrSer-----CysLeu 963
QY 1401 CTGGCCAGGGTGAAGGACAGCAGGAGGACACTGT----- 1436
Db 964 AspSerLysGlyAlaAspGlyGlyTrpGlnHisCysValTrpSerSerSerLeuGlnGln 983
QY 1437 -----TCCCCAAGCTTGCTGAGCAGATGATCCTGGAGGAGTTTGGAAAGTGCCTGTGG 1490
Db 984 CysLeuSerProSer-TyrLeu-----ProLeuArgCysMetAlaGlyGlyCysGln 1000
QY 1491 CAGATTCTCCACATCGAGTTCAGTCCAAAGGGGCTCAAAATTTAGTAGCTGCAAGCTGG 1550
Db 1000 yArgLeu-----LeuArgGlyProGluSerCysSerLeuGlyCysAlaGlnAlaPhe 1017
QY 1551 TAAA-----GGGAATGCCTGTGGCAAGCGCTCAGC 1580
Db 1017 rGlnCysAlaLeuCysLeuArgArgProHisCysGlyTrpCysAlaTrpGlyGlyGlnAs 1037
QY 1581 CCTGGGAATCTCCACGGAGGAGCTGGTCCCGAGGGA-----GG 1619
Db 1037 pGlyGlyArgCysMetGluGlyLeuSerGlyProArgAspGlyLeuThrCysGln 1057
```

```
QY 1620 AGCAGAGGCGCGCATCATGCGCAGGTCAAGTCAAGTGT 1653
Db 1057 yArgProGlyAlaSerTrpAlaPheLeuSerCys 1068

RESULT 6
T08852
Lustrin A - California red abalone
C:Species: Haliotis rufescens (California red abalone)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08852
R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E.
J. Biol. Chem. 272, 32472-32481, 1997
A:Title: Molecular cloning and characterization of lustrin A, a matrix protein from s
A:Reference number: Z16496; MUID:98070424; PMID:9405458
A:Accession: T08852
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1428 <SHE>
A:Cross-references: EMBL:AF023459; NID:q2723361; PIDN:AAB95154.1; PID:q2723362
A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial
C:Superfamily: antileukoproteinase repeat homology
C:Keywords: extracellular matrix; extracellular protein
F:1382-1426/Domain: antileukoproteinase repeat homology <ALP>
```

Alignment Scores:

Pred. No.:	0.0169	Length:	1428
Score:	131.50	Matches:	141
Percent Similarity:	29.36%	Conservative:	51
Best Local Similarity:	21.56%	Mismatches:	244
Query Match:	3.28%	Indels:	218
DB:	2	Gaps:	31

US-09-743-237-1 (1-2241) x T08852 (1-1428)

```
QY 498 GAAATCAAGAAGCAGGTGGTAGTGTG-----CCA 527
Db 297 GluIleSerCysAlaGlyGlyAlaCysProValAsnThrValCysValAlaHisPro 316
QY 528 GCGGAGCCCTGAGAGACGACGCTTTCAGGCCCTCTGGCTCAGAAATCTGTGGCAAG 587
Db 317 SerGlyAlaProAlaValCysCysPheLysPro-----AlaGlyProThrProGln 334
QY 588 TTCCCATCATCCAGAGGACGAGGAGGCTCCAGCTGCCCTCGGAAGAAGACTCCAGC 647
Db 335 ProProThrIleProGlnProProThrThrProSerSerProThrGlyAspProCysGlu 354
QY 648 CCATGGTGATTGTGCAGCTGAAAGGAGGCGCCAGATGCTCTGCATAGAC----- 698
Db 355 ProGlyValAsnValAsnCysThrAlaGlyThrCysArgLeuValValAspCysArgPhe 374
QY 699 ----AACTGTGGCGGAGGAGCTCAAGGCTCCATCTCTCTCCTCAGTACGATGACCAG 755
Db 375 ProGlyCysProAla-----ValProLysCysValAspPro 386
QY 756 AGCAGTTTCCCTCAGTCAGAGCTCCCTAAGCCCAATGACAACTTTAGTGGGAAGACTTCG 815
Db 387 SerSerLysProSer-----LeuAsnCysSerIleGlyAsp----- 398
QY 816 CCAGTACCAGCG-----AAGTTAAATCTCATCACACAGGTTGATAATGGAGCTCTC 866
Db 399 -----ProAlaLeuAsnProAsnLeuGlnGluIleSerCysValGlyAlaAlaCys 416
QY 867 CCATCAGCTGTCAATGGGGCTCCCTTCCCTCTGGA---CCTGCTCTG----- 911
Db 417 ProArgAsnThrAlaCysPheAlaAlaProSerGlySerProAlaValCysCysThr 436
QY 912 CAAGGCGCCACCAAAATAACTCTGTGGGTACTGTGACTGTCTCTCCAGCGGGGACTTC 971
Db 437 SerGlyProProArgProGluProProSerProSerPro-----ProThrGlyAspPro 454
QY 972 TGCAAC-----AGCTGCAGCTGCAACAACCTGCGC-----CAT 1004
Db 972 TGCAAC-----AGCTGCAGCTGCAACAACCTGCGC-----CAT 1004
```



```

Db 476 aglyAsnProSerLeuHisGly-::: ||| ::: |||
QY 1588 TTCCACGGCTGAGCTGGCCAGGACATCCCTTTACCAGCTTGACGCTACTCAATT 1529
Db 491 nserGln-::: ||| ::: |||
QY 1528 TTACGCCCTTGAGCTTGAACATCGATGGAGAACTCGGACAGGACGACCTTC- 1477
Db 502 eSerAlaPro-::: ||| ::: |||
QY 1476 -----CAAACCTCTCCAGGATCA--TCTGCTCAGCAAGCTTGGGAA 1436
Db 518 rArgGlnAsnLysProGlnProProAlaSerIleSerValPro- 534
QY 1435 CAGTGTCTCTGCTCTGCTCCTCACCCTGGCCAGGAGGACACATGGCCCTCCT 1376
Db 535 -----AlaProAlaPheIle-::: ||| ::: |||
QY 1375 ACTTCCAGGAGATACAGGAGGCTGCTATTTTCTCAGTTTGGAGGCTCTGAG 1316
Db 550 tValProSerLys-::: ||| ::: |||
QY 1315 AACTTGGCTGGGCAATAATAGCTGCTCTCAAAGTCCCAGGCTCCATGTAGTGGGT 1256
Db 559 -----ThrLysSerProHisLeu-::: ||| ::: |||
QY 1255 GTGCTCATCAGCATTTTCTGCTGAGCTTCTTCATAGTTTTCGAAGCAATCATTTG 1196
Db 568 uAlaSerAlaAsnPhe-::: ||| ::: |||
QY 1195 CAATGGAAGAACACATGATTTTGGCTCATAGCACTACAGTACTTCTTCAGGACG- 1138
Db 574 -----SerHisProSerSerAlaSerG1 582
QY 1137 -----CTGAGCGCTTACAGTTGACAGTTCACCTTTGCTG 1109
Db 582 uValSerSerProTyrLeuThrValIleProAsnAsnAlaTyrSerPheGlnLeuSer 602
QY 1108 TGTGGAAGTTTAGCAGCTCCAGGAGGCTTTCCTCCATTTTGGTTCGAAAGCTT- 1054
Db 602 rThiLeaArgGlyGlyThrProSerGlnAlaValProPheTyrAsnGlySerPheTyr 622
QY 1053 -CAGGATTTCTATCAAGACAGCCTTTATGCTT-::: ||| ::: |||
Db 622 rProGlnMetPheGlnGlnProGlnProGlnProLeuGlnArgGlnSerGlnAlaGlnArgG1 642
QY 1006 TCATGGCGCAGGTTGTTGACAGCTGCTGTCAGAGTCCCGCTGGAGACAGCTCA 947
Db 642 uSerLysAlaSerSerCysSerSerSerHisArgGlnPro- 656
QY 946 CAGTACCCAGACAGATTTATTTTGGTGGCCCTTGCAGAGCAGGTCCAGAGGAAAGGCA 887
Db 657 -----GlnValSerValAsnSerLeuSerSerGlnAlaAsnValGlnGlnHisArgG1 674
QY 886 GCCCATTTGACAGCTGATGGGAGAGCTCCATTATCAACCTGTGTGATGAGATTTAACTTC 827
Db 674 n----- 674
QY 826 GCTGGTACTGCCAGAACTCTTCCCACTAAAGTTGCTCATTTGGCTTAGGAGCTCTGACTGA 767
Db 675 -----MetSerGlnLysPheGluValAla-::: ||| ::: |||
QY 766 GGGAAACTGCTGCTGCTCATCTACTAGGAGAGCAGATGGAGCGCTTTGAGCTCCCTCCGCG 707
Db 687 pSerArg-::: ||| ::: |||
QY 706 CCACAGTTGCTATGCAGAGCATCTGGCGCCTCTCTTTCAGCTGACAAATCACCATGGG 647
Db 698 -----ProPheGlyGlnIleMetAlaAlaProValGlnProGlnAs 711
QY 646 CTGGAGTCTTCT-::: ||| ::: |||

```

```

Db 711 nPheSerMetSerPheAlaSerIleAlaSerSerAlaProProAlaThrLeuAsnPheSe 731
QY 609 -----CTGCCTCTCGGATGATG-::: ||| ::: |||
Db 731 rSerAsnGlyTyrHisIleSerThrProProGlyValAlaHisGlnLysAsnHisGlnSe 751
QY 591 -----GGAACCTGCAACAGGATTCCTGAGCCAGAGGCGCTGG 554
Db 751 rSerGluAlaLysThrGlyGlySerCysSer-::: ||| ::: |||
QY 553 AAAGCTCGCTCTTCAGGCG-::: ||| ::: |||
Db 769 -LysLysAsnLeuGlnGlyLysProGlnGlyMetMetAsnGlyHisThrLeuVal-PheA 788
QY 499 TCAACACTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440
Db 788 spAsnProSerArgThrLeuAsnPheValSerGlyThrTrpPro- 802
QY 439 CACCAAAATGATTTCCCTGTA-::: ||| ::: |||
Db 803 -----ProProAlaAlaThrAlaIleAsnGlyAspProSerValPheThrG 818
QY 416 --CACCTAGCAGTACAGTGTAGAGGAGGAGATCAGAGTTCCTCTAGTATATATAGC 359
Db 818 InHisLeuThrGlnArg-::: ||| ::: |||
QY 358 AATTTCAAGCAATTCGTATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGT 299
Db 836 isSerGlnAlaAspSerValSerAlaThrSerSerGlnTrpLysAsnProAlaThrSer- 855
QY 298 TCTACATAGTATTAAAGAACTGTCCCTTATCTTACAGAGGGGGAGATCATCACTGAAA 239
Db 856 -----SerSerLeuThrSerCysThrSerLeuAsnLeuLysGlnPheGlnSerGlnG 874
QY 238 AATCTCTAATATGGAAGTCTTTGGTGTGCAAGAAAAAATCTCCAGCCCAACCCAA 179
Db 874 InIleArgThrHisGlyGlnThrGlnIleSerPheAlaIaProThrAsnProGlnPro 894
QY 178 ACTACAGGCTGAAGACTATCTGCAAGAACAGCAAA--AAGGCCCTGGTTCATCCCTA 122
Db 894 er-----GlnGlyLysGlnGlyArgSerGlyGlySerSerPro 907
QY 121 GC-----CCTGCTGGGCGGAGGTGGTGGAGTCAACAA 89
Db 907 erValThrGlySerAlaSerHisGlyLysProAlaAsnSerLysValSerAsnSerLysA 927
QY 88 GT-----CCAACAGCCAGGAAACACACA 66
Db 927 laLeuLeuLeuSerProValProLeuSerGlnGluHisThr 940
RESULT 11
S50832
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C:Accession: S50832
R:Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.;
Nature Genet. 8, 177-181, 1994
A:Title: Structure and expression of the gene responsible for the triplet repeat diso
A:Reference number: S50832; MUID:95144175; PMID:7842016
A:Accession: S50832.
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1184 <NAG>
A:Cross-references: EMBL:D31840
C:Genetics:
A:Gene: GDB:DRPLA; B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p13.31-12p13.3112p12p
Alignment Scores:
Pred. No.: 0.162 Length: 1184
Score: 119.50 Matches: 125

```

Percent Similarity:	29.91%	Conservative:	44
Best Local Similarity:	22.12%	Mismatches:	206
Query Match:	2.95%	Indels:	191
DB:	2	Gaps:	24

US-09-743-237-1 (1-2241) x S50832 (1-1184)

QY	1991	C C C C T C C A A T A T C T C G T C - - - - -	-A C A G C C T G G C C T G G A A A A C C C C T G A	1944
D b	254	I I I I I : : : I I : : : I I I : : :	S e r S e r S e r V a l S e r S e r S e r A l a S e r G l y A l a P r o P h t h r	273
QY	1943	T G C C C A C C T G A C T T G T A T G T T G G C A G A G T C C A C T G A A C A G A G C C T A G C T C T C T	1884	
D b	274	L y s P r o P r o T h r T h r P r o v a l - G l y G l y A s n L e u P r o S e r A l a P r o P r o P r o A l a --	292	
QY	1883	A A C T T G A G G C T G G C T C A T A A T A A C T T G A A T G A G G T A C T A C T A A G C A A T T T G A A A C A A A	1824	
D b	292	-- -- -- -- --	292	
QY	1823	C A C A T A A C A A G A G T A G A A A T A C C C T C T T T T T A T G C A G G C A G G A G T C C C A G A G G G	1764	
D b	293	-- -- -- -- -- A s n P h e P r o H i s V a l T h r P r o A s n L e u P r o P r o P r o A l a	306	
QY	1763	C T C C C C A G G A C A G A G G G C C A G T G C C A T C C C A G G T G G C T C T C G C T T G C A C T T	1704	
D b	306	a L e u A r g P r o L e u A s n A s n A l a S e r A l a S e r P r o G l y L e u G l y A l a -- -- -- --	322	
QY	1703	T A G G A T A C C T T G A G T A G G T G G C C A G T A C C A T G C A G C A C T C A G A C C T C A C A C G -- -- --	1649	
D b	323	-- -- -- -- -- G l n P r o L e u P r o G l y H i s L e u P r o S e r P r o f r y A l a M e t G l	336	
QY	1648	-- -- -- -- -- T G A C C T G G C A T G A T G C G C G G C C T G C T C T C C C T C C C T G G G --	1610	
D b	336	y G l n G l y M e t G l y L e u P r o P r o G l y P r o G l u L y s G l y P r o T h r L e u A l a P r o S e r P r	356	
QY	1609	-C A C A G C T T C C T C G G T G C A G A T T C C A G G C T G A G G C T T C C C A C A G G C A N T T C C C T T T A	1551	
D b	356	O H i s S e r L e u P r o - P r o A l a S e r S e r A l a P r o A l a P r o P r o M e t A r g P h e P r o f r y S	376	
QY	1550	C A G C T T G C A C G C T A C T C A A T T T T C A G C C C T T G A C T T G A A C T G A T G G A G A A T C T G	1491	
D b	376	e r S e r -- -- -- -- -- S e r S e r S e r S e r S e r S e r A l a a	384	
QY	1490	G C A G A G G A C C T T C C A A A C T C C C A G G A T A C T C G T C A C C A A G C T T G G G A A C A G T G	1431	
D b	384	l a L a S e r S e r S e r S e r S e r S e r S e r S e r A l a S e r P r o P h e P r o A l a S e r G l n A	404	
QY	1430	C T C C T C C T C T G C T C T C A C C C T G G C C A G C A G G C A G C A C A T G T G C C C C A C T A C T T C	1371	
D b	404	l a L e u P r o S e r T y r P r o H i s S e r P h e P r o -- -- -- -- -- p r o P r o T h r S e r L	418	
QY	1370	C C A G G A G A T A C A G G A G A G G C C T A T T T T T T C A G T T T T G G A G G T C C T G A G A A C T T	1311	
D b	418	e U s e r V a l S e r A s n G l n P r o P o l y S t y r T h r G l n P r o S e r L e u P r o S e r G l n A l a V a l t	438	
QY	1310	G G -- -- -- -- -- C T G G G A C A A A T A A T G G C T G C T	1287	
D b	438	r p s e r G l n G l y P r o P r o P r o P r o P r o P r o T y r G l y A r g L e u L e u A l a A s n S e r A s n A l a H	458	
QY	1286	C T C A A A G T C C C A G G C C C A T G A T G A G G G G T G T C A T C A G C A N T T T T G T T C T G G A C T	1227	
D b	458	i s P r o g l y P r o P r o P r o S e r T h r G l y A l a G l n S e r T h r A l a H i s P r o P r o V a l S e r T	478	
QY	1226	T T C T T C A T A G T T T T G C A A G C A A T C A N T T G C A A N T G G A A G A C -- -- -- --	1183	
D b	478	h r H i s H i s H i s H i s G l n G l n G l n G l n G l n G l n G l n G l n G l n G l n H	498	
QY	1182	-- -- -- -- -- A C A T G A T T T G G C C C T C A T - - - - -	1164	
D b	498	i s H i s G l y A s n S e r G l y P r o P r o P r o G l y A l a P h e P r o H i s P r o L e u G l u C l y G l y S	518	
QY	1163	G C A C T A C - - - - - A G T A G T T C T T C A G G C A G C C T G A G C G C T T A C A	1125	

Db	518	erSerHisHisAlaHisProTyrAlaMetSerProSerLeuGlySerLeuArgProTyrP	538
Qy	1124	GTTC-----CACCCCTTGCTGTCGAAGTTTACGACCTC	1090
Db	538	roProGlyProAlaHisLeuProProHisSerGlnValSerTyrSerGlnAlaGlyP	558
Qy	1089	CCAGACGGCCTTTCCCACTTTTGGTTGGAAAGCTTCAGGATTCTATCAAGACACACGCCCT	1030
Db	558	roAsnGlyPro---ProVal-----	563
Qy	1029	TTATGGCTTTGAAGCCCTCGAGCTCATGCCGAGGTTGTTGCAGCTGCAGCTGTTCGAGA	970
Db	564	-----SerSerSerSer-----AsnS	569
Qy	969	AGTCCCGCTGGAGAAGCAGCTCACACTACCCAGACAGAGTTATTTGGTGGTCCCTTGC-	911
Db	569	erSerSerSerThrSerGlnGlySerTyr-----ProCysS	581
Qy	910	-----AGAGCAGCTCCAGAGGAAAGGCAGCCCCCATTTGACAGCTGATGGGA	865
Db	581	erHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPhe-----	596
Qy	864	GAGCTCCATTATCAACTGTGTGATGAGATTTAACTTCGCTGGTACTGGCAGAGAAGTCTTC	805
Db	597	--ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValIle	613
Qy	804	CCACTAAAGTTGTCATTGGCTTAGGAGCTCTGACTGAGGGAAACTGCTCTGGTCATCGT	745
Db	614	-----AlaThrValAlaSerSerProAlaGlyTyrLys	624
Qy	744	ACTGAGGAAGCAGATGGACGCTTTGAGCTCCCTCGCCACAGTTGTCTATGCAGAGCA	685
Db	625	ThrAlaSerProGly-----ProProProTyrGlyLysArgAlaProSer	640
Qy	684	TCTGGGGCCCTCTTCAGCTGACAATCACCATCGGGCTGGAGTCTTTCTTCGAGAGGC	625
Db	641	ProGlyAlaTyrIysThrAla-----ThrProProGlyTyrLysProGlySerProPro	658
Qy	624	AGCTGGAGGCCCTCTGCCTCTGGATGATGGGAACCTTGCACACAGGATTCCTTGAGCCA	565
Db	659	SerPheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProPro	674
Qy	564	GAGGGCTGGAAAGCTGCTCTTCAGGCGTCCG-----CCTGGCACACTACCAC	514
Db	675	AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuProP	694
Qy	513	CTGCT	509
Db	694	roAla	695

RESULT 12

S09257 homeotic protein Hox A4 - chicken
N:Alternate names: homeotic protein Chox1-4
C:Species: Gallus gallus (chicken)
C:Date: 29-Jan-1993 #sequence_revision
C:Date: 29-Jan-1993 #text_change 21-Jul-2000
C:Accession: S09257; S10883
R:Sasaki, H.; Yokoyama, E.; Kuroiwa, A.
Nucleic Acids Res. 18, 1739-1747, 1990
A:Title: Specific DNA binding of the two chicken deformed family homeodomai
A:Reference number: S09256; MUID:90245562; PMID:1970866
A:Accession: S09257
A:Molecule type: mRNA
A:Residues: 1-309 <SAS>
A:Cross-references: EMBL:X52670; NID:963218; PIDN:CAA36896.1; PID:963219
R:Scotting, P.J.; Hewitt, M.; Keynes, R.J.
Nucleic Acids Res. 18, 3999, 1990
A:Title: Isolation and analysis of chick homeobox cDNA clones.
A:Reference number: S10883; MUID:90326535; PMID:1973835
A:Accession: S10883
A:Molecule type: mRNA
A:Residues: 207-273 <SCO>
A:Cross-references: EMBL:X52747; NID:963223; PIDN:CAB57949.1; PID:96018426

C;Function:
A;Description: control of embryonic development by tissue- and stage-specific regulation
C;Superfamily: homeotic protein Hox D4; homeobox homology
C;Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation
F;210-266/Domain: homeobox homology <Hox>

Alignment Scores:					
Pred. No.:	0.214	Length:	309		
Score:	117.50	Matches:	58		
Percent Similarity:	36.3%	Conservative:	14		
Best Local Similarity:	29.2%	Mismatches:	81		
Query Match:	2.90%	Indels:	45		
DB:	1	Gaps:	9		

US-09-743-237-1 (1-2241) x S09257 (1-309)

1829	QY	AGCAAAACATACACAAAGTAGAAANAATCAACCCCTCTTTTATGACGGCAGGAGTCCCC	1770
			1771
24	Db	ThrGlnHisSerGlySerAlaGlySerAlaSerTyHisProHisHisProHisPro	43
1769	QY	AGAGGCGCTCCCAAGGACAGGAGGCCAGTGTCCATCCCAAGGTGCCTGCTGCTGCTT	1710
			1711
44	Db	HisAlaProProPro-----Pro- ProProProProHisLe	57
1709	QY	GCACCTTAGGATACCTTGAGTAGGCTGGCCAGTACCATGCAGATCACTCAGACCTCACAG	1650
			1651
57	Db	uHisAlaAlaHisPro-----GlyProAlaLeuProGluTyPheProArgProArgAr	75
1649	QY	CTGACCTGCCCATGATGGCGGCTCTGCTCCTCCCTGGGCACACAGTCTCCTCGGTGCAG	1590
			1591
75	Db	gGluProGlyTyGlnAla---ProAlaAlaProProGlyProProGlyProPro-----	92
1589	QY	ATTCCAGCGGCTGAGGCTTGCCACAGGCAATCCCTTTACACAGCTTCACGCGTACTCAAT	1530
			1531
93	Db	---ProGluAlaLeuTyProAlaGlnAlaProSerTyProGlnAlaProTySer-T	111
1529	QY	TTTCAGCCCTTGGACTTGAACCTCGATGTGGAGAACTGCGACAGGCACTTCCAAACCTC	1470
			1471
111	Db	yrSerSer-----AlaGlySerAlaAlaProGlyProGluGlnProp	125
1469	QY	CTCCAGGATCATCTCTCAGCCCAAGCTGGGGACAGTGTCTCTGCTCTGCTCTCCTCAACC	1410
			1411
125	Db	roProGlyAlaSerProPro-----ProProAlaLysGlyHisP	140
1409	QY	CTGGGCCACGACGAGGCACATGTGGCTCTCCACTACTTCCC-----AGGA	1365
			1366
140	Db	roGlyProAla-----GlnProLeuLeuProGlyHisAlaLeuGlnArga	155
1364	QY	GATACAGGAGAAGCGCTGCCTATTTTTCTCAGTTTTTGGAGGCTCTGAGAACCTGGCTGG	1305
			1306
155	Db	rgcysGluAlaAlaProAla-----AlaGlyAlaG	165
1304	QY	GGACAAATANTGGCTGCTCTCAAGTCCCAAGGCTCCATGTAGTGGGTG	1255
			1256
165	Db	lyThrGlyProGlycysAlaLeuLeuProAspLysSerLeuProGlyLeu	181

RESULT 13

Tl13954
 MEGF6 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C:Accession: Tl13954
 C:R:R.Nakayama, M.; Nakajima, M.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 Genomics 51, 27-34, 1998
 A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Accession: Tl13954
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1574 <NAK>
 A:Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294
 A:Experimental source: strain Sprague-Dawley; brain
 C:Genetics:


```

Db 302 GlyGlyGlnValPheCysLysAspSerCysSerCys----- 313
QY 1014 CGTCTCAAGCCATAAAGGGGTGCTT-----GATAGAAATCCTGAAGCTTTCCAAACCA 1067
Db 314 -----ValCysProGlyGlyAspLysLysThrCysThrAlaPro 327
QY 1068 AAAATGGGAAAGCCGCTCTGGAGCTGCTAAACTTCGACACACAAAGGGTGCAACTGT 1127
Db 328 GlnValTyrAspGly-----ValAlaCysSerCys 337
QY 1128 AAG-----CGCTCAGGCTGCTG----- 1145
Db 338 SerCysProValAsnMetGlnLysProAlaAspCysProArgProGlnLysTyrPasp 357
QY 1146 AAGAACTACTGT-----GAGTGCTATGAGGCCAAATCATG 1181
Db 358 LysGluCysArgCysGluCysProValLysHisAspCysLysAsnGlyLysValTrp 377
QY 1182 TGTCTTCCATTGCAATCATTTGCT----- 1208
Db 378 AspGluThrIleCysGlnCysIleCysProArgAspAlaProValCysThrAlaGlyLys 397
QY 1209 -----TCAAAAATCTAGAGAAATCTCAGAACCAAAATG 1244
Db 398 GluArgCysGlyGluSerCysGluCysLysCysIleAsnArgGluProLys-GluGlyCy 417
QY 1245 CTGATGAGCACACCCACTCATGAGCTGGGAGCTTTCGAGAGCAGCCATTATTGTTC 1304
Db 417 s-----AlaLysProLeuValTrpAsnGluAsnThrCysLysCys-----ValCysPr 433
QY 1305 CCAGCCAAAGTTCAGAGACCTCCAAACTCAGAAAAAATAGCAGGCGCTTCTCCTGTATC 1364
Db 433 o-----AlaAspLysGln-----Metse 439
QY 1365 TCCTGGG----- 1371
Db 439 rProGlyGlyCysGlySerGlyLysSerPheAsnLysLeuThrCysGlnCysGluCysAs 459
QY 1372 -----AAGTAGTGGAGGCCACATGTGCTCGCTCGCTGGCC 1406
Db 459 pGlnSerAlaSerLysCysGlyLeuLysArgTrpAsnAlaAspThrCysLysCys----- 477
QY 1407 CAGGTGAGGAAGCAGACGAGGAGCACTGTTCGCCAAGCTTGGCTGAGCAGATGATCCTG 1466
Db 478 -----GluCysGlnProGlyMetProGluGlyCysGlyLysGlnThrTr 493
QY 1467 GAGGAGTTTGAAGGTCCTTCGCACATTCCTCCACATCGAGTTCAGTCCAAAGGGGCTG 1526
Db 493 pIleSerAspLysCysLysCysGlu-CysSerPro---ThrIleThrCysGlnAlaProG 512
QY 1527 AA-----AATTGAGTAGCGTGCAAG-----CTGTAAGGGGA 1559
Db 512 InIleLeuAspLeuAsnThrCysGluCysLysCysProValAsnMetLeuAlaGlnLysG 532
QY 1560 ATGCTGTGCAAGCCTCAGCCCTGGGAATCTGCACGAGGAAGCTGGTGC----- 1610
Db 532 LuLysCysLysSerProArgGlnTrpThrAspSer---LysCysLeu-CysGluCysSer 550
QY 1611 -----CCAGGGAGGACAGAGCGCGGCATCATGCCAGGTAG 1649
Db 551 ThrThrProAlaThrCysGluGlyLysGlnThrTrpCysGlyGlu 565
```

RESULT 15

UIHU

thyroglobulin precursor, major splice form - human

C:Species: Homo sapiens (man)

C:Date: 28-Feb-1986 #sequence_revision 05-Nov-1999 #text_change 16-Jun-2000

C:Accession: A59110; S00014; A01532; S03422; I38343; I57669; S02266; S39432; S66241; S62

R:Walthery, Y.

submitted to the EMBL Data Library, April 1988

A:Description: Human mRNA for thyroglobulin.

A:Reference number: A59110

A:Accession: A59110
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2767 <MAL1>
A:Cross-references: GB:X05615; NID:g37173; PIDN:CAA29104.1; PID:g37174
A:Note: revision to S00014
R:Walthery, Y.; Lissitzky, S.
Eur. J. Biochem. 165, 491-498, 1987
A:Title: Primary structure of human thyroglobulin deduced from the sequence of its 84
A:Reference number: S00014; MUID:87246630; PMID:3595599
A:Accession: S00014
A:Molecule type: mRNA
A:Residues: 1-1041, Y, 1043-1057, T, 1059-2767 <MAL2>
A:Cross-references: GB:X05615; NID:g37173
A:Note: this sequence is revised in A59110
R:Walthery, Y.; Lissitzky, S.
Eur. J. Biochem. 147, 53-58, 1985
A:Title: Sequence of the 5'-end quarter of the human-thyroglobulin messenger ribonucleic acid
A:Reference number: A01532; MUID:85127024; PMID:3971976
A:Accession: A01532
A:Molecule type: mRNA
A:Residues: 1-730 <MAW>
A:Cross-references: GB:X02154; NID:g37175; PIDN:CAA26089.1; PID:g1335349
A:Note: the translated sequence in GenBank entry HSTYRRK5, release 111.0, (PIDN:CAA26089.1)
R:Parma, J.; Christophe, D.; Pohl, V.; Vassart, G.
J. Mol. Biol. 196, 769-779, 1987
A:Title: Structural organization of the 5' region of the thyroglobulin gene. Evidence
A:Reference number: S03422; MUID:88062712; PMID:3681978
A:Accession: S03422
A:Molecule type: DNA
A:Residues: 1-134, 'Q', 136-415; 640-652, 'G', 654-733, 'A', 735-737; 880-983, 'DR', 985-999 <P>
A:Cross-references: EMBL:X06059; NID:g37145; PIDN:CAA29454.1; PID:g1359884; EMBL:X06059
R:Christophe, D.; Cabrer, B.; Bacolla, A.; Targovnik, H.; Pohl, V.; Vassart, G.
Nucleic Acids Res. 13, 5127-5144, 1985
A:Title: An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream
A:Reference number: I38343; MUID:85269632; PMID:2991855
A:Accession: I38343
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-22, 'GKF' <CHR>
A:Cross-references: EMBL:X02749; NID:g37162; PIDN:CAA26527.1; PID:g758106
R:Targovnik, H.M.; Cochaux, P.; Corach, D.; Vassart, G.
Mol. Cell. Endocrinol. 84, R23-R26, 1992
A:Title: Identification of a minor Tg mRNA transcript in RNA from normal and goitrous
A:Reference number: I57669; MUID:92347597; PMID:1639210
A:Accession: I57669
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1503-1508, 'L', 1567-1601 <REW>
A:Cross-references: GB:S40807; NID:g252170; PIDN:AAB22685.1; PID:g252171
A:Note: this sequence fragment represents a minor splice form
R:Marrig, C.; Lejeune, P.J.; Venot, N.; Vinet, L.
FEBS Lett. 242, 414-418, 1989
A:Title: Hormone synthesis in human thyroglobulin: possible cleavage of the polypeptide
A:Reference number: S02266; MUID:89121111; PMID:2914619
A:Contents: evidence for Tyr-149 as ring donor for thyroxine formation
A:Accession: S02266
A:Molecule type: protein
A:Residues: 101-109; 114-121; 126-131; 143-149 <MAR>
A:Note: only the first peptide was sequenced; others were isolated and their amino acid
R:Gentile, F.; Salvatore, G.
Eur. J. Biochem. 218, 603-621, 1993
A:Title: Preferential sites of proteolytic cleavage of bovine, human and rat thyroglobulin
A:Reference number: S39431; MUID:94094855; PMID:8269951
A:Accession: S39431
A:Molecule type: protein
A:Residues: 20-27; 522-527; 541-547; 616-619, 'X', 621; 1001-1005; 1009-1011, 'X', 1013; 1424-1426
R:Xiao, S.; Pollock, H.G.; Taurog, A.; Rawitch, A.B.
Arch. Biochem. Biophys. 320, 96-105, 1995
A:Title: Characterization of hormonogenic sites in an N-terminal, cyanogen bromide fragment
A:Reference number: S66241; MUID:95314327; PMID:7793989
A:Accession: S66241

Db 1503 AlapheSerGlnThrHisCysValThrAspCysGlnArgAsnGluAlaGlyLeuGlnCys 1522
QY 1308 -----GCCAAGTTCTCAGGACCTCCAAATAGAGAAAAATAGGACGCTTCTCC 1358
Db 1523 AspGlnAsnGlyGlnTyrArgAlaSerGlnLysAspArgGlySerGlyLysAlaPhe--- 1541
QY 1359 TGTATCTCC-----TGGAGTAGTCTGGAGGCC-----ACA 1388
Db 1542 CysValAspGlyGluGlyArgLeuProTyrTrpGluThrGluAlaProLeuGluAsp 1561
QY 1389 TGTGCTGCTGCTGCTGCCAGGCTGAGAA----- 1418
Db 1562 SerGlnCysLeuMetMetGlnLysPheGluLysValProGluSerLysValIlePheAsp 1581
QY 1419 -----GCAGACGAGGAGCACTGTTCCCAAGCTTGGCTGAGCAGATGATC 1463
Db 1582 AlaAsnAlaProValAla-ValArgSerLysValProAspSerGluPhePro-----Va 1599
QY 1464 CTGGAGAGTTTG-----GAAGTGCTGTCGCGAGATTCACACATCGAG 1508
Db 1599 lMetGlnCysLeuThrAspCysThrGluAspGluAlaCysSerPhePheThrValSerTh 1619
QY 1509 TTCAAGTCCAAAGGGCTGAAATGAGTACGCTGCAAGCTGTAAGGGGAATCC----- 1563
Db 1619 rThrGluPro-----GluIleSerCysAspPh 1628
QY 1564 ----CTGTGCAAGCCTCAGCCCTGGGAATCTGCAC-----GAGGAAGCTGGTGCCCA 1613
Db 1628 eTyrAlaTrpThrSerAspAsnValAlaCysMetThrSerAspGlnLysArgAspAlaLe 1648
QY 1614 GCGAGGACGAGCGCGCCGATCGCCAGGTCAGCTGTGAGTGTGATGAT-CTGCA 1672
Db 1648 uGlyAsnSerLysAlaThrSerPheGlySerLeuArgCysGlnValLysValArgSerHl 1668
QY 1673 TGGTACTGGCCAGCCTACT-----CAAGGTATCTCTAAAGTGCACGCA 1714
Db 1668 sGlyGlnAspSerProAlaValTyrLeuLysLysGlyGlnGlySerThrThrLeuGl 1688
QY 1715 GGCAGAG----CCACCCTGGGGATGGACACTGGCCCTCCTGCTGCGGGAGGCC----- 1766
Db 1688 nLysArgPheGluProThrGlyPheGlnAsnMetLeuSerGlyLeuTyrAsnProIleVa 1708
QY 1767 -----TCTGGGACTCCTCGCCC-----TGCATAAAAAGAGGGTG 1801
Db 1708 lPheSerAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuAlaCysAs 1728
QY 1802 ATTTTCTACTTGTGTATGTTGCTTTCAAATTCCTTAGTAGTACCTCCATTCAAGT 1861
Db 1728 pArgAspLeuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleIleCy 1748
QY 1862 T---ATTATGAGCCAGCTCAAGTTAGAGAGTAGGCTCTTCTCAGGTGGAGCTGCCCC 1918
Db 1748 sGlyLeuLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspPr 1767
QY 1919 AAATCATACATAAGTCAGTCGCCATCAGGGGTTTTTCCAGGCCAGCCCTGTGACAGGAG 1978
Db 1767 oSerGlu-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyrAspGlnGl 1784
QY 1979 ATATGGAGGGGGCTGGGTTAGAGTGGGT 2009
Db 1784 u----SerHisGlnValIleLeuArgLeuGly 1793

Search completed: April 21, 2003, 11:43:07
Job time : 118.924 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:19:57 ; Search time 30.0475 Seconds

(without alignments)
6186.760 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tctctcgtgggtgcccgcg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q/cn2_1/USPTO_spool/US09743237/runat_21042003_111943_388/app_query.fasta_1.4494
-DB-SwissProt_40 -QFMT-fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -CGEN_1_1_93 @runat_21042003_111943_388 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	39.9	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	31.2	299	1 MTL5_HUMAN	Q9v4i5 homo sapien
3	144	3.6	1183	1 DRPL_RAT	P54258 ratus norv
4	125.5	3.1	892	1 HIC1_MOUSE	Q9riy5 mus musculus
5	119.5	3.0	1185	1 DRPL_HUMAN	P54259 homo sapien
6	117.5	2.9	309	1 HXAA_CHICK	P17277 gallus gall
7	117.5	2.9	971	1 RECK_HUMAN	O95980 homo sapien
8	117.5	2.9	1700	1 BAR3_CHITE	O03376 chironomus
9	115.5	2.9	1509	1 GSR1_HUMAN	Q9nzm4 homo sapien
10	113	2.8	643	1 CD93_RAT	Q9et61 ratus norv
11	112.5	2.8	323	1 AMFR_HUMAN	P26442 homo sapien
12	112.5	2.8	2768	1 THYG_HUMAN	P01266 homo sapien
13	112	2.8	652	1 CD93_HUMAN	Q9npy3 homo sapien
14	111.5	2.8	946	1 IP3L_HUMAN	P27987 homo sapien
15	110.5	2.8	493	1 OC90_HUMAN	Q02509 homo sapien
16	109.5	2.7	478	1 EGR4_HUMAN	Q00911 ratus norv
17	108.5	2.7	610	1 LEM2_HUMAN	P16581 homo sapien
18	107.5	2.7	1093	1 SM5B_MOUSE	Q60519 mus musculus

C 19	107	2.6	1142	1	MGC1_HUMAN	O60732 homo sapien
C 20	106.5	2.6	1436	1	WC11_BOVIN	P30205 bos taurus
C 21	106.5	2.7	2805	1	MAPA_HUMAN	P78559 homo sapien
C 22	106	2.6	470	1	EGR2_MOUSE	P08152 mus musculus
C 23	105.5	2.6	611	1	LEM2_CANFA	P33730 canis famil
C 24	104.5	2.6	545	1	AIRE_HUMAN	O43918 homo sapien
C 25	104.5	2.6	760	1	OCT1_XENLA	P16143 xenopus lae
C 26	104.5	2.6	3149	1	TGCU_EBV	P03186 epstein-bar
C 27	104	2.6	1093	1	AFI7_HUMAN	P55198 homo sapien
C 28	104	2.6	1096	1	KPCL_ASPNG	Q00078 aspergillus
C 29	104	2.6	1680	1	FUR2_DROME	P30432 drosophila
C 30	104	2.6	1724	1	POV_HUMAN	Q9ukk3 homo sapien
C 31	104	2.6	1964	1	NTC4_MOUSE	P13695 mus musculus
C 32	104	2.6	5376	1	ZAN_MOUSE	O88799 mus musculus
C 33	103	2.6	579	1	SOC7_MOUSE	Q8vnd2 mus musculus
C 34	103	2.6	1139	1	KPCL_TRIRE	Q99014 trichoderma
C 35	103	2.5	1801	1	LMB2_RAT	P15800 ratus norv
C 36	102	2.5	470	1	EGR2_RAT	P51774 ratus norv
C 37	102	2.5	867	1	SSPO_BOVIN	P98167 bos taurus
C 38	102	2.5	2768	1	THYG_RAT	P08882 ratus norv
C 39	102	2.5	5376	1	ZAN_MOUSE	O88799 mus musculus
C 40	101.5	2.5	3828	1	TRX_DROVI	Q24742 drosophila
C 41	101.5	2.5	4660	1	LRP2_RAT	P98158 ratus norv
C 42	101	2.5	503	1	IRF7_HUMAN	Q92985 homo sapien
C 43	101	2.5	1509	1	GSR1_HUMAN	Q9nzm4 homo sapien
C 44	101	2.5	1581	1	LMB3_MOUSE	Q9t0b6 mus musculus
C 45	101	2.5	1816	1	LMA4_HUMAN	Q16363 homo sapien

ALIGNMENTS

RESULT 1

MTL5_MOUSE

ID MTL5_MOUSE STANDARD; PRT; 295 AA.

AC Q9WTJ6;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific

DE Metallothionein-like protein).

GN MTL5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ICR; TISSUE=Testis;

RX MEDLINE=99208669; PubMed=10191092;

RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;

RT "A novel testis-specific metallothionein-like protein, tesmin, is an

RT early marker of male germ cell differentiation.";

RL Genomics 57:130-136(1999).

CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.

CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES

CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED

CC PROGRESSIVELY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U77383; AAD24667.1; -

CC EMBL; U67176; AAD24666.1; -

CC MGD; MGI:1340029; Mtl5.

CC InterPro; IPR005172; CXC.

CC Pfam; PF03638; CXC; 1.

CC Spermatoogenesis.

DR

DR

DR

KW

FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 9 86e-117 Length: 295
Score: 1599.00 Matches: 295
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.90% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-1 (1-2241) x MTL5_MOUSE (1-295)

```
QY 651 ATGGTATTTTCAGCTCAAGAGGAGGCGCCAGATGCTCTGCATAGACAACGTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 711 AGGAGCTCAAGCGCTCCATCTCTCTCAGTACGATGACGACGAGTTCCTCCCTCAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerSerPheProGln 40
QY 771 TCAGAGCTCCCTAAGCAATCACAACTTTAGTGGGAAGACTTCTGCCAGTACCAAGCAAG 830
DB 41 SerGluLeuProLysProMetThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 831 TTAATCTCATCACACAGGTTGATATGGAGCTCTCCCATCAGCTGCTCAATGGGCGTCC 890
DB 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 891 TTTCCCTCTGAGCTCTGCTGCAAGGCGCCACCAAAATCACTCTGTCTGGGTACTGTGAC 950
DB 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTGCGCATGAGCTC 1010
DB 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnLeuArgHisGluLeu 120
QY 1011 GAGCGCTTCAAGCCATAAAGCGCTGCTTGTATGAGAATCCTGAAGCTTTCCCAACCAAAA 1070
DB 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
QY 1071 ATGGGGAAGCGCTCTGGGAGCTGCTTAACCTCGACACAGCAAGGTTGCAACTGTAAAG 1130
DB 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
QY 1131 CGCTCAGCGCTCCCTGAAGAACTACTGTGAGTGCTATGAGGCGCAAAATCATGTGTTCTTC 1190
DB 161 ArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIleMetCysSerSer 180
QY 1191 ATTTGCAATGCTGCTTGCAGAACTATGAGAAAGTCCAGAACGAAATAATGCTGATG 1250
DB 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
QY 1251 AGCACCCCACTACATGGAGCCCTGGGAGCTTTGAGAGCAGCCCAATATTGTCCTCCAGCC 1310
DB 201 SerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuSerProAla 220
QY 1311 AAGTTCTCAGGACCTCCAAAACCTGAGAAAATAAGGAGCGCTTCTCTCTGATCTCTCTGG 1370
DB 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTip 240
QY 1371 GAAGTAGTGAGGCCCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
DB 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGluAlaGluGlnGlu 260
QY 1431 CACTGTTCCCAAGCTGGCTGAGCAGCATGATCTCTGAGGAGGTTTGAAGTGCTGCTGCTG 1490
DB 261 HisCysSerProSerLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSer 280
QY 1491 CAGATTCTCCACATCGAGTTCAAGTCCCAAGGGGCTGAAAATTGAG 1535
DB 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
```

RESULT 2
ID MTL5_HUMAN STANDARD; PRT; 299 AA.
AC Q9Y4I5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL: U86074; AAD24668.1;
DR Genes: HGNC:7446; MTL5.
DR MIM: 604374;
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 1.
DR KEGG: Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;
Alignment Scores:
Pred. No.: 1 2e-89 Length: 299
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 31.21% Indels: 4
DB: 1 Gaps: 1
US-09-743-237-1 (1-2241) x MTL5_HUMAN (1-299)
QY 651 ATGGTATTTTCAGCTCAAGAGGAGGCGCCAGATGCTCTGCATAGACAACGTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 711 AGGAGCTCAAGCGCTCCATCTCTCTCAGTACGATGACGACGAGTTCCTCCCTCAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGln 40
QY 771 TCAGAGCTCCCTAAGCAATCACAACTTTAGTGGGAAGACTTCTGCCAGTACCAAGCAAG 830
DB 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 831 TTAATCTCATCACACAGGTTGATATGGAGCTCTCCCATCAGCTGCTCAATGGGCGTCC 890
DB 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
QY 891 TTTCCCTCTGAGCTCTGCTGCAAGGCGCCACCAAAATAACTCTGTCTGGGTACTGTGAC 950
DB 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTGCGCATGAGCTC 998

QY 1663 -----CTCAGACCT-----CACAGCTGACCTGCCATGATGGCGGCC 1626
|||||
Db 305 oProAlaLeuArgProLeuAsnAlaSerAlaSerProGlyMetGlyAlaGlnPr 325
QY 1625 TCTG-----CTCCTCCC 1614
|||
Db 325 oileProGlyHisLeuProSerProHisAlaMetGlyGlnGlyMetSerGlyLeuProPr 345
QY 1613 TGGG-----CACAGCTTCTCGGTGCAGATT 1587
|||||
Db 345 oGlyProGlyLysGlyProThrLeuAlaProSerProHisProLeuProProAlaSerS 365
QY 1586 CCCAGGCTGAGGCTGCCAGCAGCATCCCTTTACAGCTTGCACGCTACTCAATTTT 1527
|||||
Db 365 erSerAlaProGlyProProMetArgProTyProSerSerCysSerSerSerVala 385
QY 1526 CAGCCCTTGGACTTGAACCTGATGGAGAACTGCGACAGGACCTTCCAACTCCT- 1468
:||||
Db 385 laAla-----SerSerSerSerAlaAlaThrSerGlnTyrProA 399
QY 1468 ----- 1468
Db 399 laSerGlnThrLeuProSerTyrProHisSerPheProProProThrSerMetSerValS 419
QY 1467 -----CCAGGATCATCTGCTCAGCCAACTTGGGGAACAGTGTCT- 1429
|||||
Db 419 erAsnGlnProProLysTyrThrGlnProSerLeuProSerGlnAlaValTyrSerGlnG 439
QY 1428 -----CCTGCTGTGCTTCT- 1407
|||||
Db 439 lyProProProProProProProTyroGlyArgLeuLeuProAsnAsnAsnThHisProG 459
QY 1406 GGCA-----GCAGGAGGACACATGTGGCTCCACTACTTCCAGGAGATACA 1359
|||||
Db 459 lyProPheProProThrGlyGlyGlnSerThrAlaHisProProAlaProAlaHisHis 479
QY 1358 GGAGAGGCTGCCTATTTTCTCAGTTTGGAGGCTCTGAGAACTTGGCTGGGACAA 1299
:|:
Db 479 isHisGlnGln----- 482
QY 1298 ATATGCTGCTCTCAAGTCCCG- 1273
|||
Db 483 -----GlnGlnGlnProGlnProGlnProGlnProGlnHisHisGlyA 499
QY 1272 -----GCTCCATGATGAGTGGGTGCTCATCAGCATTTTTCGTTCTGGACTTTCTTC 1221
|||
Db 499 snSerGlyProProProProGlyAla----- 507
QY 1220 ATAGTTTTTGAACCAATGCAATTTGCAATGGAGAACACATGATTTTGGCTCATAGCA 1161
:|||||
Db 508 -----TyrProHisProL 512
QY 1160 CTCACACTAGTCTTCA- 1137
|||||
Db 512 euGluSerSerAsnSerHisHisAlaHisProTyrAsnMetSerProSerLeuGlySerL 532
QY 1136 TGAGCGCTTACAGTTG-----CACCCCTTGTCTGTGTCGAA 1102
|||
Db 532 euArgProTyrProProGlyProAlaHisLeuProProSerHisGlyGlnValSerTyrS 552
QY 1101 GTTTAGAGCTCCAGAGCGCTTTCCCATTTTGGTTGGAAGCTTCAGGATTTCTAT 1042
|||
Db 552 erGlnAlaGlyProAsnGlyProProValSerSerSerSerAsnSerSerGlySerS 572
QY 1041 CAACACAGCGCTTTATGCTTTGAAGCTGCGAGCTCATGCCAGCTTGTTCAGCTGC 982
:|||||
Db 572 erGlnAlaAlaTyrSerCysSerHisProSerSer----- 584
QY 981 AGCTGTTGCAGAGATCCCGCTGGAGAGCAGTCACAGTACCCACAGACAGATTTTGG 922
Db 584 ----- 584

QY 921 GTGGCCCTTCAGAGCAGGTCCAGAGGAAAGCAGCCCATTCACAGCTGATGGGAGAG 862
|||||
Db 585 -----GlnGlyProGlnGlyAlaSerTyrProPhe-----P 595
QY 861 CTCATATCAACTGTGTGATGAGATTAACTTCGCTGCTACTGGCAGAGTCTTCCCA 802
|||||
Db 595 roProVal-ProProfile-----ThrThrSerSerAlaThrLeuSerThrValile--- 611
QY 801 CTAAGTGTCAATTGGCTTAGGGAGCTCTCACTGAGGGAACCTGCTGCTGCTCATCTACT 742
|||
Db 612 -----AlaThrValaSerSerProAlaGlyTyrLysThr 623
QY 741 GAGGAAGCAGATGAGCGCTTTGAGCTCCTCGCGCCACAGTTGTCTATGACAGCATCT 682
|||
Db 624 AlaSerProGly-----ProProGlnTyrSerLysArgAlaProSerPro 639
QY 581 GGGCGCTCTCTTCAGCTGACAAATCACCATGGGCTGGAGTCTTTCTCCGAGGCGCAGC 622
|||
Db 640 GlySerTyrLysThrAla-----ThrProGlyTyrLysProGlySerProProSer 657
QY 621 TGAGCGCTCTCTGCTCTGCGATGATGGAACTTGCAACAGGATTCCTGAGCCAGAG 562
:||||
Db 658 PheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProProAla 673
QY 561 GGGCGTGAAGTGGCTCTTCAGGCTGCGC-----CCTGGCACACTACCAGCTG 511
|||||
Db 674 GlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuProProA 693
QY 510 CT-----TCTTTGATTTCACACTT 491
|||
Db 693 laglyProSerSerLeuSerSerLeu 701
RESULT 4
HICL_MOUSE STANDARD; PRT; 892 AA.
ID HICL_MOUSE Q9R1Y5; Q9R1Y6; Q9R2B0;
AC Q9R1Y5; Q9R1Y6; Q9R2B0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypermethylated in cancer 1 protein (Hic-1).
GN HIC1.
OS Mus musculus (Mouse).
OC Eukaryota; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC STRAIN=129/Sv, and Swiss Webster; TISSUE=Embryo;
RX MEDLINE=99172081; PubMed=10072440;
RA Grimm C., Spoerle R., Schmid T.E., Adler I.-D., Adamski J.,
Schughart K., Graw J.;
RT "Isolation and embryonic expression of the novel mouse gene Hic1, the
RT homologue of Hic1, a candidate gene for the Miller-Dieker syndrome.";
RL Hum. Mol. Genet. 8:697-710(1999).
RN [2]
RP SEQUENCE FROM N.A., DISEASE, AND TISSUE SPECIFICITY.
RX MEDLINE=20122251; PubMed=10655551;
RA Carter M.G., Johns M.A., Zeng X., Zhou L., Zink M.C., Mankowski J.L.,
Donovan D.M., Baylin S.B.;
RT "Mice deficient in the candidate tumor suppressor gene Hic1 exhibit
RT developmental defects of structures affected in the Miller-Dieker
RT syndrome.";
RL Hum. Mol. Genet. 9:413-419(2000).
RN [3]
RP SEQUENCE OF 179-338 FROM N.A.
RC STRAIN=129/Sv; TISSUE=Liver;
RX MEDLINE=99297610; PubMed=10371200;
RA Guerardel C., Deltour S., Leprince D.;
RT "Evolutionary divergence in the broad complex, trantrack and bric a
RT brac/poxviruses and zinc finger domain from the candidate tumor
RT suppressor gene hypermethylated in cancer.";
RL FEBS Lett. 451:253-256(1999).
CC -!- FUNCTION: Putative transcription factor. May act as a tumor

suppressor. May be involved in development of head, face, limbs
 and ventral body wall.
 -I- SUBUNIT: Interacts with CtBP (By similarity).
 -I- SUBCELLULAR LOCATION: Nuclear (Probable).
 -I- ALTERNATIVE PRODUCTS: A number of isoforms may be produced by
 alternative splicing.
 -I- TISSUE SPECIFICITY: Ubiquitously expressed with highest levels in
 heart and lung.
 -I- DEVELOPMENTAL STAGE: Expression is first detected in the embryo
 after 9 days post coitum. In the embryo, expression is found in
 restricted regions of somite derivatives, limb anlagen and cranio-
 facial mesenchyme. In the fetus, it is additionally expressed in
 mesenchymes apposed to precartilaginous condensations, at many
 interfaces to budding epithelia of inner organs, and weakly in
 muscles.
 -I- DISEASE: Defects in HIC1 are the cause of perinatal death with
 serious developmental anomalies, including craniofacial, exencephaly,
 cleft palate, omphalocele, craniofacial and limb anomalies.
 -I- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
 -I- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 -I- SIMILARITY: CONTAINS 5 C2H2-TYPE ZINC FINGERS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.
FT DOMAIN 704 707 POLY-PRO.
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 94 MISSING (IN REF. 2).
FT CONFLICT 333 333 Y -> H (IN REF. 2).
FT CONFLICT 339 339 M -> I (IN REF. 2).
FT CONFLICT 541 541 P -> T (IN REF. 3).
FT CONFLICT 1028 1028 G -> A (IN REF. 2).
SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

Alignment Scores:
Pred. No.: 0.188 Length: 1185
Score: 119.50 Matches: 125
Percent Similarity: 29.91% Conservative: 44
Best Local Similarity: 22.12% Mismatches: 206
Query Match: 2.95% Indels: 191
DB: 1 Gaps: 24

US-09-743-237-1 (1-2241) x DRPL_HUMAN (1-1185)

QY 1991 CCCCCTCCCATATCTCTGTC-----ACAGGCGTGGCCTGGRAAAACCCCTGA 1944
DB 254 ProProProThrThrProLeuSerValSerSerGlyAlaSerGlyAlaProProThr 273
QY 1943 TGCCACCTGACTGTGTATGTATGGGAGAGTCCACCTGAAGAAGAGAGCTAGCTCTCT 1884
DB 274 LysProProThrThrProVal-GlyGlyAsnLeuProSerAlaProProAla-- 292
QY 1883 AACTTGAGGCTGCTCATATACTTGAATGGAGGTACTACTAAGCAATTTGAAGCAAA 1824
DB 292 ----- 292
QY 1823 CACATAACAACAGTAGAAATCACCTCTTTTATGACGGGCGAGGAGTCCCGAGGG 1764
DB 293 -----AsnPheProHisValThrProAsnLeuProProProAl 306
QY 1763 CTTCCCGAGGACAGAGGCGCAGTGTCCATCCCGAGGTGGCTCTGCTGCTTGCACCT 1704
DB 306 aLeuArgProLeuAsnAlaSerAlaSerProProGlyLeuGlyAla----- 322
QY 1703 TAGGATACCTTGAGTAGGCTGGCCAGTACGATCAGATCAGACCTCAGCAGC----- 1649
DB 323 -----GlnProLeuProGlyHisLeuProSerProTyAlaMetG1 336
QY 1648 -----TGACCTGGCCATGATGCGGCGCTCTGCTCTCCCTCGG-- 1610
DB 336 yGlnGlyMetGlyGlyLeuProProGlyProGlyLysGlyProThrLeuAlaProSerPr 356
QY 1609 -CACCAGCTTCTCGGTGCAGATTCACAGGCTGAGGCTGGCCAGGATTCCTCCCTTA 1551
DB 356 oHisSerLeuPro-ProAlaSerSerAlaProAlaProAlaProMetArgPheProTyr 376
QY 1550 CCAGCTTGACGCTACTCAATTTTACGCCCTTGGACTTGAACCTGATGTGGAGATCTG 1491
DB 376 erSer-----SerSerSerSerAla 384
QY 1490 CGACAGGCACTTCCAAACTCTCCAGGATCATCTGCTCAGCCCAAGCTGGGGAACAGTG 1431
DB 384 laAlaSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnA 404
QY 1430 CTCTGCTGCTGCTCTCCAGGCTGGCCAGGAGGACATGTGGCCTCCACTACTTC 1371
DB 404 laLeuProSerTyProHisSerPhePro-----ProProThrSerL 418
QY 1370 CCAGGAGATACAGGAGGCGCTGCTATTTTCTCAGTTTGGAGGCTCTGAGAACTT 1311
DB 1311 -----

DB 418 euSerValSerAsnGlnProProLysTyThrGlnProSerLeuProSerGlnAlaValr 438
QY 1310 GG-----CTGGGCAAAATAATGGCTGCT 1287
DB 438 rpSerGlnGlyProProProProProProProTyGlyArgLeuAlaAsnSerAsnAlah 458
QY 1286 CTCAAAGTCCCAAGGCTCATGTAGTGGGTGTGCTCAGCATTTTCTGTTCTGCACT 1227
DB 458 isProGlyProPheProProSerThrGlyAlaGlnSerThrAlaHisProProValSerT 478
QY 1226 TTTCTTCATAGTTTTCGAAGCAATGCAATTTGCAATTTGGAAGAAC----- 1183
DB 478 hrHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 498
QY 1182 -----ACATGATTTTGGCCTCAT-----A 1164
DB 498 isHisGlyAsnSerGlyProProProProGlyAlaPheProHisProLeuGlnGlyCys 518
QY 1163 GCACTCAC-----AGTAGTTCTTCAGGAGCTGAGGCTTACA 1125
DB 518 erSerHisHisAlaHisProTyAlaMetSerProSerLeuGlySerLeuArgProTyP 538
QY 1124 GTTG-----CACCTTTTGTGTCTGCTGCAAGTTTACAGCTC 1090
DB 538 roProGlyProAlaHisLeuProProProHisSerGlnValSerTySerGlnAlaGlyP 558
QY 1089 CCAGAGCGGCTTTCCCATTTTGGTTGGAAGCTTCAGGATTTCTATCAAGACAGCCT 1030
DB 558 roAsnGlyPro-----ProVal----- 563
QY 1029 TTATGCTTTGAAGCGCTCAGCTCATGCGCAGGTGTTGACAGCTGACGCTGTCGAGA 970
DB 564 -----SerSerSerSer-----AsnS 569
QY 969 AGTCCCGCTGGAGAAGACAGTACAGTACCACAGAGATTATTTGGTGGCCCTTGC- 911
DB 569 erSerSerSerThrSerGlnGlySerTy-----ProCys 581
QY 910 -----ACAGCAGGTCCAGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 865
DB 581 erHisProSerProSerGlnGlyProGlnGlyAlaProTyProPhe----- 596
QY 864 GAGCTCCATATCAACCTGTGTGATGAGATTTAACTTCTGCTGCTGCTGCTGCTGCTGCTGCT 805
DB 597 --ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValIle 613
QY 804 CCACTAAAGTTGTCATGGCTTAGGAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 745
DB 614 -----AlaThrValAlaSerSerProAlaGlyTyLys 624
QY 744 ACTGAGGAGCAGATGAGGCGCTTTGAGCTCCCTCGCCGACAGATGCTCTATCAGAGCA 685
DB 625 ThrAlaSerProProGly-----ProProProTyGlyLysArgAlaProSer 640
QY 684 TCTGGGCGCTCTCTTTCAGCTGACAATCACCATGGGCTGGAGTCTTCTTCTCCGAGGCG 625
DB 641 ProGlyAlaTyLysThrAla-----ThrProGlyTyLysProGlySerProPro 658
QY 624 AGCTGAGGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565
DB 659 SerPheArgThrGlyThrProProGlyTyArgGlyThr-----SerProPro 674
QY 564 GAGGCGCTGGAAGCTGCTTTCAGGCTGCGC-----CCTGGCACACTACCAC 514
DB 675 AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuPro 694
QY 513 CTGCT 509
DB 694 roAla 695

RESULT 6
HXAA4_CHICK PRT: 309 AA.
ID HXA4_CHICK

CC -1- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
 CC by suppressing MMP-9 secretion and by direct inhibition of its
 CC enzymatic activity. RECK down-regulation by oncogenic signals may
 CC facilitate tumor invasion and metastasis. Appears to also
 CC regulate MMP-2 and MT1-MMP, which are involved in cancer
 CC progression.
 CC -1- SUBUNIT: Interacts with MMP-9.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
 CC cells. It is undetectable in tumor-derived cell lines and
 CC oncogenically transformed cells.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D50406; BAA34060.1; -
 CC EMBL: AL158830; CAD13384.1; -
 CC Genew; HGNC: 11345; RECK.
 CC MIM: 605227; -
 CC HSP; P80424; IAN1.
 CC InterPro: IPR002350; kazal.
 CC SMART: SM00280; KAZAL; 3.
 CC SMART: SM00011; VWC_def; 1.
 CC PROSITE: PS00282; KAZAL; 1.
 CC Signal; Glycoprotein; GPI-anchor; Serine protease inhibitor;
 KW Membrane; Anti-oncogene; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 942 REVERSION-INDUCING CYSTEINE-RICH PROTEIN
 FT WITH KAZAL MOTIFS.

FT PROPEP 943 971 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 942 942 GPI-ANCHOR (POTENTIAL).
 FT DOMAIN 632 677 KAZAL-LIKE 1.
 FT DOMAIN 708 750 KAZAL-LIKE 2 (DEGENERATE).
 FT DOMAIN 753 787 KAZAL-LIKE 3 (DEGENERATE).
 FT DOMAIN 37 338 5 X KNOT REPEATS.
 FT REPEAT 37 84 KNOT 1.
 FT REPEAT 104 141 KNOT 2.
 FT REPEAT 151 197 KNOT 3.
 FT REPEAT 216 263 KNOT 4.
 FT REPEAT 292 338 KNOT 5.
 FT DISULFID 635 654 BY SIMILARITY.
 FT DISULFID 633 677 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 971 AA; 106456 MW; 173D47D6AE6F834 CRC64;

Alignment Scores:
 Pred. No.: 0-257 Length: 971
 Score: 117-50 Matches: 84
 Percent Similarity: 30.84% Conservative: 52
 Best Local Similarity: 19.05% Mismatches: 141
 Query Match: 2.93% Indels: 164
 DB: 1 Gaps: 23

US-09-743-237-1 (1-2241) x RECK_HUMAN (1-971)

QY 660 TGTACGTGAAGAGCGCGCCAGATCCTCTGCATACACAACTGTGGCGGAGGAGCTC 719
 Db 422 CysHisSerLysSerArgGlySerIleCysLysSerAspCys----- 436
 QY 720 AAGCGCTCCATCTGCTTCTCTAGTAGGACGACAGAGAGTTCCTCAGTCAGAGCTC 779
 Db 437 -----ValGluIleLeuLysLysCysGlyAspGlnAsnLysPheProGluAspHisThr 454

QY 780 CCTAAGCCCAATGACAACTTTAGTGGGA-----AGACTTCTGCCAGTA 821
 Db 455 AlaGluSerIleCysGluLeuLeuSerProThrAspLeuLysAsnCysIleProLeu 474
 QY 822 -----CCAGCGAAGTTA---AATCTCATCACACAGGTTGATAATGAGCT 863
 Db 475 AspThrTyrlleuArgProSerThrLeuGlyAsnIleValGluGluValThrHisProCys 494
 QY 864 CTCCCATCA-----GCTGTCAATGGGCTGCTTTCCTCT 899
 Db 495 AsnProAsnProCysProAlaAsnGluLeuCysGluValAsnArgLysGlyCysProSer 514
 QY 900 GGA---CCTGCTCTGAAGGCCACCAATAACTCTGCTGGTACTGACTGCTTC 956
 Db 515 GlyAspProCysLeu-----ProTyrlleuValGlnGly---CysLysLeuGly 530
 QY 957 TCCAGGGGGGACATCTGCAACAGCTGCAGCTGCAACACCTGCCCATGAGCTCGAGCGC 1016
 Db 531 GluAlaSerAspPhe----- 535
 QY 1017 TTCAAAGCCATAAAGGGTCTCTTGTATAGAAATCTCGAACCTTTCACCAACCAAAATGGGG 1076
 Db 536 -----IleValArgGlnGlyThrLeuIleGlnValProSerSer 548
 QY 1077 AAAGCGCTCTGGGAGCTGCTAAACTTCGACACAGCAAGGGTGCACACTCTAAAGCGCTCA 1136
 Db 549 AlaGlyGluValGlyCysTyrlleu-----CysSerCysGlyGlnSer 563
 QY 1137 GGCTGCTGAAG----- 1148
 Db 564 GlyLeuLeuGluAsnCysMetGluMetHisCysIleAspLeuGlnLysSerCysIleVal 583
 QY 1149 -----AACTACTGTGAGTGC 1163
 Db 584 GlyGlyLysArgLysSerHisGlyThrSerPheSerIleAspCysAsnValCysSerCys 603
 QY 1164 TATGAGCGCAAAATCATGCTGCTTCTTC----- 1190
 Db 604 PheAlaGlyAsnLeuValCysSerThrArgLeuCysLeuSerGluHisSerSerGluAsp 623
 QY 1191 -----ATTGCAAAATGCATT-----GCTTGC 1211
 Db 624 AspArgArgThrPheThrGlyLeuProCysAsnCysAlaAspGlnPheValProValCys 643
 QY 1212 -----AAAACTATGAAGAAAGTCCAGACGAAATGCTGATGACACACCC 1259
 Db 644 GlyGlnAsnGlyArgThrTyrlleuValCysIleAlaArgCysValGlyLeuGlnAsp 663
 QY 1260 CACTACATGAGCGCTGGGAGCTTTGAGAGCAGCCATTATTTGTCGCCGCAAGTTCTCA 1319
 Db 664 HisGlnPheGluPheGlySerCysMetSerLys-----AspProCys----- 677
 QY 1320 GGACCTCCAAACTGAGAAAAATAGCAGGCC----- 1352
 Db 678 AsnProAsnProCysGlnLysAsnGlnArgCysIleProLysProGlnValCysLeuThr 697
 QY 1353 -----TTCCTCTGTATCTCTCGGAAGTAGTGAGGCCACATGCTGCTGCTG 1400
 Db 698 ThrPheAspLysPheGlyCysSerGlnTyrlleuValProArgGlnLeuAlaCysAsp 717
 QY 1401 CTGGCCCGAGGCTGAGGAAGCAGACGAGCAGCTGTTTCCCA-----AGCTTG 1448
 Db 718 GlnValGlnAspProValCysAspThrAspHisMetGluHisAsnAsnLeuCysThrLeu 737
 QY 1449 GCTGAGCAGATGATCTCGAGGAGTTTGAAGGTGCTGTCGAGAGTTCTCCATCGAG 1508
 Db 738 TyrGln-ArgGlyLysSerLeuSerTyrlleuValGlyProCysGlnProPheCysArgAla 757
 QY 1509 TTCA-----AGTCCAAAGGGGTGAAATTTAGTAGCTGCAAGCTGGT----- 1551
 Db 757 rGluProValCysGlyHisAsnGlyLeuThrTyrlleuValCysAlaAlaTyrlleu 777

Qy 1407 CAGGCTGAGGAAGCAGAGCAGTGTTCCTCCCAAGCTTGCTGAGCAGATGATCTGT 1466
 Db 478 -----GlucyGlnProGlyMetProProGlyGlyLysGlnThrTr 493
 Qy 1467 GAGGAGTTTGAAGTCTGCTGAGATTCCTCACATCGAGTTCACAGTCCCAAGGCTG 1526
 Db 493 pIleSerAspLysCysLysGlu-CysSerPro---ThrIleThrCysGlnAlaProG 512
 Qy 1527 AA-----AATTGAGTACGCTGCAAG-----CTGTAAGGGGA 1559
 Db 512 inIleLeuAsnThrCysGluCysLysCysProValAsnMetLeuAlaGlnLysG 532
 Qy 1560 ATGCTGTGGCAAGCTCAGCCCTGGGAATCTGCACGAGGAAGCTGTGC----- 1610
 Db 532 LuLysCysLysSerProArgGlnThrAspSer---LysCysLeu-CysGluCysSer 550
 Qy 1611 -----CGAGGAGGAGCAGAGCGCGGCATCATGCGCAGGTGAG 1649
 Db 551 ThrThrProAlaThrCysGluGlyLysGlnThrTrpCysGlyGlu 565

RESULT 9

GSRL_HUMAN STANDARD; PRT; 1509 AA.
 AC Q9NZM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glioma tumor suppressor candidate region gene 1 protein.
 GN GLTSCR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=20175430; PubMed=10708517;
 RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
 RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
 RA Scheichauer B.W., Louis D.N., Jenkins R.B.;
 FT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
 RT region.";
 RL Genomics 64:44-50(2000).
 CC -1- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
 CC placenta, skeletal muscle, and pancreas, and at lower levels in
 CC lung, liver, and kidney.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF182077; AAF62874.1;
 CC Genew; HGNC:4332; GLTSCR1.
 DR MIM; 605690;
 DR DOMAIN 37 45 POLY-GLY.
 FT DOMAIN 884 889 POLY-PRO.
 FT DOMAIN 1214 1225 POLY-SER.
 FT DOMAIN 1282 1286 POLY-PRO.
 FT DOMAIN 1294 1304 POLY-PRO.
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

Alignment Scores:

Pred. No.: 0.404 Length: 1509
 Score: 115.50 Matches: 140
 Percent Similarity: 32.03% Conservative: 56
 Best Local Similarity: 22.88% Mismatches: 217
 Query Match: 199
 DB: 1 Gaps: 29

US-09-743-237-1 (1-2241) x GSRL_HUMAN (1-1509)
 Qy 2065 GTTTTCTCACTTCAAGCAGAAATACAGGAAGAAAAACGCAAAAATCCAAACACCC 2006
 Db 783 IlePheValIleGlnAsnGlnLeuGlyValProProAlaSerAsnPro---AlaPro 801
 Qy 2005 AGCTCTCAACCGAGCCCTCCCATATCTCTGTCCAGGCTGCTGGAAAAACCCCT 1946
 Db 802 ThrAlaProGlyProProGlnProLeuArgProGlnSerGlnPro-----Pro 818
 Qy 1945 GATGGCCACCTGACTTGTATGTGATTTGGGAGAGTCCACCTGAAGAGACCCTAGCTCT 1886
 Db 819 GluGlyProLeu-----ProProAlaProHisLeuProProSerSer 832
 Qy 1885 CTAACCTTGAGCTGCTCATATACTTGAATGAGGTACTACTAAGCAATTTGAAGCA 1826
 Db 833 ThrSer-----SerAlaValAlaSerSerSerGluThr 843
 Qy 1825 AACACATAACAACAGTAGAAATACCCCTCTTTTATGCAGGCGAGGAGTCCCCAGAG 1766
 Db 844 SerSer-----ArgLeuProAla 849
 Qy 1765 GGCCTCCCGAGGACAGAGGCGCAGTGTCCATCCCGAGGTGGCTGTGCTGTGTCAC 1706
 Db 850 ProThrProSerAspPheGlnLeuGlnPheProProSerGlnGly----- 864
 Qy 1705 TTTAGGATACCTTGCAGTAGGCTGCCAGTACCAGTCCAGATCAGATCAGACCT- 1655
 Db 865 -----ProHisLysSerProProProProProProThr 874
 Qy 1654 ---CACAGCTCACCTGGCCATGATCGCGGCTCTGTCTCCCTCCCT---GGCACCAGCTT 1601
 Db 875 LeuHisLeuValProGluProAlaAlaProProProProProProAlaThrPheGlnMet 894
 Qy 1600 CCTCGTGCGAGTTCCCGAGGCTG-----AGG 1574
 Db 895 ValThrThrProPheProAlaLeuProGlnProLysAlaLeuLeuGluArgPheHisGln 914
 Qy 1573 CTTGCCACAGCAGCTCCCTTTTACCAGCTTGCAGCTACTCAATTTTTCAGCCCTTGAC 1514
 Db 915 ValProSerGlyIleIleLeuGlnAsnLysAla-GlyGlyAlaProAlaProGlnTh 934
 Qy 1513 TTGACCTCGATGTGAGAAATCTGCGACAGGACCTTCCAAACTCTCTCAGGATCATGCG 1454
 Db 934 rSerThrSerLeuGlyPro-----LeuThrSerProAlaAlaSerVa 948
 Qy 1453 T-----CAGCCAAGCTTGGGAGACAGTGTCTCTGTCTCTCCTCACCC 1409
 Db 948 lLeuValSerGlyGlnAlaProSerGlyThrProThrAlaProSerHisAlaProAlaPr 968
 Qy 1408 TGGGCCACGAGGAGGACATGTGCTCCACTTCTCCAGGAGATACAGGAGGCC 1349
 Db 968 oAlaProMetAlaAlaThrGlyLeuProProLeuProAlaGluAsnLysAla----- 986
 Qy 1348 TGCCTATTCTCTCAGTTTGGAGTCTCGAAGCTTGGCTGGGAGCAATAATGGCTG 1289
 Db 987 -----PheAlaSerAsnLeuProThrLeuAsnValAlaLysAlaAlaSerSerGly-- 1003
 Qy 1288 CTCTCAAAGTCCCGAGGCTCCATGTAGTGGGTGTGCTATCAGCATTTTTCGTTCTGGA 1229
 Db 1004 -----ProGlyLysPro---SerGlyLeuGlnThrGluSerLysLeu--SerGly 1018
 Qy 1228 CTT-----TCTTCATAGTCTTTTCCAGCAGCAATGCAT 1199
 Db 1019 LeuLysLysProProThrLeuGlnProSerLysGluAlaLysPheLeuGlu-----His 1036
 Qy 1198 TTGCAAAATGGAAGAACACATGTTTGGCTCTATAGCATCTCACAGTAGTTC-----TTC 1145
 Db 1037 LeuHisLysHisGlnGlySerValLeuHisProAspThrLysThrAlaPheProSerPhe 1056
 Qy 1144 AGCAGGCTGAGCGCTTACAGTTGCACCCCTTTGCTGTGTCGAAGTTTTCAGCAGTCCCA 1085
 Db 1057-GluAsp-----AlaLeuHisArgLeuLeuProThrHisValGlnGlyAla 1072

```

QY 1084 CGGCTTTCCCAATTTTGGTTGGAAGCTTCAGGATTTCTATCAGACACGCGCTTATG 1025
DB 1073 LeuProSerProSerAspTyrHisLysValAspGluPheGluThrValSerThrGln 1092
QY 1024 GCATTGAAGCGCTCCAGCTATCGCGGAGGTGTTGTCAGCTGCGAGTGGTCGACAGTCC 965
DB 1093 LeuLeuLysArgThrGlnAlaMetLeuAsnLysTyrArgLeuLeuLeuGluGluSer 1112
QY 965 ----- 965
DB 1113 ArgArgValSerProSerAlaGluMetValMetIleAspArgMetPheIleGlnGluGlu 1132
QY 964 -----CGCTGGAGAACGAG-----TCACAGTACCACAGACAGGATGATT--- 926
DB 1133 LysThrThrLeuAlaLeuAspLysGlnLeuAlaLysGluLysProAspGluTyrValSer 1152
QY 925 -----TTGGTGGCCCTTCGACAGCAGGTCCAGAGGAAGGAGGAGCCCATG 878
DB 1153 SerSerArgSerLeuGlyLeuProIleAlaAlaSerSerGluGlyHisArgLeuPro--- 1171
QY 877 ACAGCTGATGGGAGAGCTCCATATCAACCTGTGTGATGAGATTAACTTCGCTGGTACT 818
DB 1172 -----GlyHisGlyProLeuSerSer-----SerAlaPro 1181
QY 817 GCAGAGT-----CTTCCCACTAAAGTTGTCATT-----GGC 785
DB 1182 GlyAlaSerThrGlnProProHisLeuProThrLysLeuValIleArgHisGlyGly 1201
QY 784 TTAGGAGCTCTGACTGAGGAACCTCTGCTGCTCATCTGCTACTGAGGACGATGGAGC 725
DB 1202 AlaGlyLysSerPro-----SerValThrTrpAla-----Arg 1212
QY 724 GCTTTGAGTCCCTCGCGCACAGTGTCTATCGACAGCATCTGGCGCTCCTCTTCAGC 665
DB 1213 AlaSerSerSerLeuSerSerSerSerSerSerSerAlaAlaSerSerLeuAspAl 1232
QY 664 TGAC-----AA 659
DB 1232 aspGluAspGlyProMetProSerArgAsnArgProProIleLysThrTyrGluAla 1252
QY 658 ATCCATCGGCGCTGGAGTCTT----- 637
DB 1252 gSerArgIleGlyLeuLysLeuLysLysGlnGluAlaGlyLeuSerLysValValHi 1272
QY 636 -----TCTCCGAGGCGAGCTGGAGCGCTCTCTGCTCTGCTGATGAGGAACTTG 584
DB 1272 sAsnThrAlaLeuAspProValHisGlnProProProProAlaThrLeuLysValAl 1292
QY 583 CAACAGGATCTTCGAGCCAGAGGGGCT 556
DB 1292 aGluProProArgProProProPro 1301

RESULT 10
CD93.RAT
ID CD93.RAT STANDARD; PRT: 643 AA.
AC Q9ET61; Q9J126;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
DE subcomponent, receptor 1) (ClqR) (ClqR(p)) (C1q/MBL/SPA receptor)
DE (CD93 antigen) (Cell surface antigen AA4).
GN CLQR1 OR CD93 OR CIQRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-PVG; TISSUE=Natural killer cells;
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;

```

```

RT "Characterization and molecular cloning of rat ClqR, a receptor on NK
RT cells.";
RL Eur. J. Immunol. 30:3355-3362(2000).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=Wistar; TISSUE=Lung;
RC MEDLINE=20507883; PubMed=10934210;
RA Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;
RT "Molecular and cellular properties of the rat AA4 antigen, a C-type
RT lectin-like receptor with structural homology to thrombomodulin.";
RL J. Biol. Chem. 275:34382-34392(2000).
CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: type I membrane protein.
CC -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and
CC heart. Expressed at lower level in brain, thymus, liver, spleen,
CC intestine, kidney, adrenal gland, muscle and testis. Expressed on
CC endothelial cells, platelets, undifferentiated monocytes and
CC circulating natural killer cells.
CC -!- PTM: N- and O-glycosylated (By similarity).
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF136537; AAG01572.1; -
DB EMBL: AF160978; AAF80402.1; -
DB HSP: F35555; IEMN.
DB InterPro: IPR000152; Asx_hydroxyl.
DB InterPro: IPR000561; EGF-like.
DB InterPro: IPR001881; EGF_Ca.
DB InterPro: IPR001304; Lectin_C.
DB Pfam: PF00008; EGF_5.
DB Pfam: PF00059; Lectin_c; 1.
DB SMART: SM00034; CLECT; 1.
DB SMART: SM00181; EGF; 5.
DB SMART: SM00179; EGF_CA; 5.
DB SMART: SM00001; EGF_Like; 2.
DB PROSITE: PS00010; ASX_HYDROXYL; 3.
DB PROSITE: PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
DB PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
DB PROSITE: PS01186; EGF_2; 3.
DB PROSITE: PS01187; EGF_CA; 3.
DB Receptor; EGF-like domain; Signal; Transmembrane; Glycoprotein;
KW Repeat; Lectin.
FT SIGNAL 1
FT CHAIN 23
FT DOMAIN 24 643
FT TRANSMEM 572
FT DOMAIN 593 643
FT DOMAIN 31 173
FT DOMAIN 257 298
FT DOMAIN 299 341
FT DOMAIN 342 381
FT DOMAIN 382 423
FT DOMAIN 424 462
FT DISULFID 261 272
FT DISULFID 268 282
FT DISULFID 284 297
FT DISULFID 303 314
FT DISULFID 308 325
FT DISULFID 327 340
FT DISULFID 346 355
FT DISULFID 351 364

```

QY	1313	GTTCAGGACCTCCAAACTGAGAAAAAATAGGCAGGCGCTTCTCCTGTATCTCTCTGG--	1370
DB	360	-----GlyGlyPheHisCysGluCysTyrPva	368
QY	1371	-----GAAGTAGTGGAGGCCACA--	1388
DB	368	lGlyTyrGlnSerSerGlySerLysGluGluAlaCysGluAspValAspGlyCysThrAl	388
QY	1389	-----TCGTGCGCTG	1396
DB	388	aAlaTyrSerProCysAlaGlnGlyCysThrAsnThrAspGlySerPheTyrCysSerCy	408
QY	1397	C-----CTGCTGTGCGCCAGGGTGAGGAACGAGCAGGAGCAGCTGTCCCAAGCTT	1447
DB	408	sLysGluGlyTyrIleMetSerGlyGluAspSerThrGln	421
QY	1448	GGCTGACAGATGATCCTGGAGGAGTTTGGAAAGTGCCCTGTGCGAGATTCCTCCACATCGA	1507
DB	422	-----CysGluAspLeuGluAspGlyLeuGluAsn	431
QY	1508	GTTCAGTCCAAAGGGCTGAAATTTCAGTAGCGTGCACAGCTGCTAAAGGGAATGCCTGT	1567
DB	432	-----ProCysAspThrLeuCysIleAsnThrAspGlySerPheArgCysGlyCysProAl	450
QY	1568	GGCA--AGCCTCAGCGCTGGGAATCTGCAC-----CGAGGAAGCTGG-----TG	1609
DB	450	aGlyPheGluLeuAlaProAsnGlyValSerCysThrArgGlySerMetPheSerGluLe	470
QY	1610	CCCAGGAGG-----AGCAGAGCGCCGCGATCA-----	1637
DB	470	uProAlaArgProGlnLysGluAspLysGlyAspGlyLysGluSerThrValProLe	490
QY	1638	-----TGCCAGGTACGCTGTGAGCTGTGATGATGTCATGCTGCTACTGCCAGCCTAC	1690
DB	490	uThrGluMetProGlySerLeuAsnGlySerLysAspValSerAsnArgAlaGlnThrTh	510
QY	1691	T	1691
DB	510	T	510
RESULT 11			
ID	AMFR_HUMAN	STANDARD;	PRT; 323 AA.
AC	P26442;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Autocrine motility factor receptor precursor (AMF receptor) (GP78).		
GN	AMFR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=91302382; PubMed=1649192;		
RA	Watanabe H., Carmi P., Hogan V., Raz T., Silletti S., Nabi I.R.,		
RA	Raz A.;		
RT	"Purification of human tumor cell autocrine motility factor and		
RT	molecular cloning of its receptor.";		
RL	J. Biol. Chem. 266:13442-13448(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Placenta;		
RC	MEDLINE=95352090; PubMed=7626106;		
RA	Huang B., Xie Y., Raz A.;		
RT	"Identification of an upstream region that controls the transcription		
RT	of the human autocrine motility factor receptor.";		
RL	Biochem. Biophys. Res. Commun. 212:727-742(1995).		
CC	-!- FUNCTION: SPECIFIC RECEPTOR FOR THE AUTOCRINE MOTILITY FACTOR.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- PTM: PHOSPHORYLATED IN THE PRESENCE OF AMF		

RT ribonucleic acid and of its deduced amino-acid sequence.";
RL Eur. J. Biochem. 147:53-58(1985).
RN [5]
RP SEQUENCE OF 1-415; 640-737 AND 880-1000 FROM N.A.
RX MEDLINE=88062712; PubMed=3681978;
RA Parma J., Christophe D., Pohl V., Vassart G.;
RT "Structural organization of the 5' region of the thyroglobulin gene.
RT Evidence for intron loss and 'exonization' during evolution.";
RL J. Mol. Biol. 196:769-779(1987).
RN [6]
RP SEQUENCE OF 1-25 FROM N.A.
RX MEDLINE=85269632; PubMed=2991855;
RA Christophe D., Cabrer B., Bacolla A., Targovnik H.M., Pohl V.,
RT Vassart G.;
RT "An unusually long poly(purine)-poly(pyrimidine) sequence is located
RT upstream from the human thyroglobulin gene.";
RL Nucleic Acids Res. 13:5127-5144(1985).
RN [7]
RP SEQUENCE OF 1002-1566 FROM N.A.
RX PubMed=11124863;
RA Moya C.M., Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 5' region of the human thyroglobulin
RT gene.";
RL Eur. J. Endocrinol. 143:789-798(2000).
RN [8]
RP SEQUENCE OF 1645-2768 FROM N.A.
RX MEDLINE=99452200; PubMed=10524569;
RA Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 3' region of the human thyroglobulin
RT gene.";
RL Thyroid 9:903-912(1999).
RN [9]
RP PRESENCE OF A 11TH TYROGLOBULIN TYPE-I REPEAT.
RX MEDLINE=96390872; PubMed=8797845;
RA Molina F., Bouanani M., Pau B., Granier C.;
RT "Characterization of the type-I repeat from thyroglobulin, a
RT cysteine-rich module found in proteins from different families.";
RL Eur. J. Biochem. 240:125-133(1996).
RN [10]
RP IODINATION SITES.
RX MEDLINE=89340430; PubMed=2760035;
RA Lamas L., Anderson P.C., Fox J.W., Dunn J.T.;
RT "Consensus sequences for early iodination and hormonogenesis in human
RT thyroglobulin.";
RL J. Biol. Chem. 264:13541-13545(1989).
RN [11]
RP SULFATION.
RX MEDLINE=99382264; PubMed=10448091;
RA Niend M.-C., Cauvi D., Venot N., Chabaud O.;
RT "Sulfated tyrosines of thyroglobulin are involved in thyroid hormone
RT synthesis.";
RL Biochem. Biophys. Res. Commun. 262:193-197(1999).
RN [12]
RP VARIANT GOITER HIS-870.
RX MEDLINE=93164775; PubMed=8094490;
RA Corral J., Martin C., Perez R., Sanchez I., Mories M.T.,
RT San Millan J.L., Miralles J.M., Gonzalez-Sarmiento R.;
RT "Thyroglobulin gene point mutation associated with non-endemic simple
RT goitre.";
RL Lancet 341:462-464(1993).
RN [13]
RP VARIANTS GOITER R-1264 AND S-1996, AND VARIANTS.
RX MEDLINE=99213884; PubMed=10199792;
RA Hishinuma A., Takamatsu J., Ohshima Y., Yokozawa T., Kuma K.,
RA Yoshida S., Matsura N., Ieiri T.;
RT "Two novel cysteine substitutions (C1263R and C1995S) of thyroglobulin
RT cause a defect in intracellular transport of thyroglobulin in
RT patients with congenital goiter and the variant type of adenomatous
RT goiter.";
RL J. Clin. Endocrinol. Metab. 84:1438-1444(1999).
CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
CC (T4) AND TRIIODOTHYRONINE (T3).
CC -1- SUBUNIT: HOMODIMER.
CC PIR: A01532; UIHU.
CC PIR: S00014; S00014.

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.
CC -1- PTM: SULFATED.
CC -1- DISEASE: DEFECTS IN TG ARE A CAUSE OF FORMS OF GOITER, AN
CC ENLARGEMENT OF THE THYROID GLAND. THIS IS SOMETIMES LINKED WITH
CC HYPOTHYROIDISM.
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLTERASE/LIPASE FAMILY.
CC -1- SIMILARITY: CONTAINS 11 TYROGLOBULIN TYPE-I DOMAINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05615; CAA29104.1; -
CC EMBL; U93033; AAC51924.1; -
CC EMBL; AF230667; AAF62513.2; -
CC EMBL; AF235100; AAG39877.1; -
CC EMBL; AF230666; AAF62515.1; -
CC EMBL; AF305872; AAG17932.1; -
CC EMBL; X02154; CAA26089.1; -
CC EMBL; X06059; CAA29454.1; -
CC EMBL; X06060; CAA29454.1; JOINED.
CC EMBL; X06061; CAA29454.1; JOINED.
CC EMBL; X06062; CAA29454.1; JOINED.
CC EMBL; X06063; CAA29454.1; JOINED.
CC EMBL; X06064; CAA29454.1; JOINED.
CC EMBL; X06065; CAA29454.1; JOINED.
CC EMBL; X06066; CAA29454.1; JOINED.
CC EMBL; X06067; CAA29455.1; -
CC EMBL; X06068; CAA29455.1; JOINED.
CC EMBL; X06069; CAA29456.1; -
CC EMBL; X06870; CAA29456.1; JOINED.
CC EMBL; X02749; CAA26527.1; -
CC EMBL; AF170489; AAD51647.1; -
CC EMBL; AF170486; AAD51647.1; JOINED.
CC EMBL; AF170487; AAD51647.1; JOINED.
CC EMBL; AF170488; AAD51647.1; JOINED.
CC EMBL; AF105687; AAC95473.1; -
CC EMBL; AF105681; AAC95473.1; JOINED.
CC EMBL; AF105682; AAC95473.1; JOINED.
CC EMBL; AF105683; AAC95473.1; JOINED.
CC EMBL; AF105684; AAC95473.1; JOINED.
CC EMBL; AF105685; AAC95473.1; JOINED.
CC EMBL; AF105686; AAC95473.1; JOINED.
CC EMBL; AF080484; AAD50912.2; -
CC EMBL; AF169654; AAD50912.2; JOINED.
CC EMBL; AF169655; AAD50912.2; JOINED.
CC EMBL; AF169656; AAD50912.2; JOINED.
CC EMBL; AF169657; AAD50912.2; JOINED.
CC EMBL; AF169658; AAD50912.2; JOINED.
CC EMBL; AF169659; AAD50912.2; JOINED.
CC EMBL; AF169661; AAD50912.2; JOINED.
CC EMBL; AF169662; AAD50912.2; JOINED.
CC EMBL; AF169663; AAD50912.2; JOINED.
CC EMBL; AF169664; AAD50912.2; JOINED.
CC EMBL; AF080472; AAD50912.2; JOINED.
CC EMBL; AF080473; AAD50912.2; JOINED.
CC EMBL; AF080474; AAD50912.2; JOINED.
CC EMBL; AF080475; AAD50912.2; JOINED.
CC EMBL; AF080476; AAD50912.2; JOINED.
CC EMBL; AF080477; AAD50912.2; JOINED.
CC EMBL; AF080478; AAD50912.2; JOINED.
CC EMBL; AF080479; AAD50912.2; JOINED.
CC EMBL; AF080480; AAD50912.2; JOINED.
CC EMBL; AF080481; AAD50912.2; JOINED.
CC EMBL; AF080482; AAD50912.2; JOINED.
CC EMBL; AF080483; AAD50912.2; JOINED.
CC PIR: A01532; UIHU.
CC PIR: S00014; S00014.

Db 1691 gPhelupProThrGlyPheGlnAsnMetLeuSerGlyLeuTyraAsnProIleValPheSe 1711
 Qy 1767 ----TCYGGGACTCCCTGCC-----TGCAATAAAGAGGGTGAATTTTC 1807
 Db 1711 rAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuAlaCysAspArgAs 1731
 Qy 1808 TACTTGTGTATGTTGCTTCTCAATTCCTAGTAGTACCTCCCACTCAAGTT---AT 1864
 Db 1731. pleuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleIleCysGlyLe 1751
 Qy 1865 TATGAGCAGCCTCAAGTAGAGCTAGGCTCTTCTTTCAGGTGAGCTCTCCCAATCA 1924
 Db 1751 uLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspProSerG1 1770
 Qy 1925 CATAACAGTCAGGTGGCCATCAGGGGTTTTCAGGCAGGCCTGTGACAGAGATATGG 1984
 Db 1770 u-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyrAspGlnGlu---Se 1786
 Qy 1985 GAGGGGGTGGGTAGAGCTGGGT 2009
 Db 1786 rHisGlnValIleLeuArgLeuGly 1794

RESULT 13
 CD93_HUMAN
 ID CD93_HUMAN STANDARD; PRT: 652 AA.
 AC Q9NPY3; O00274;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement component C1q receptor precursor (Complement component 1, q
 DE subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
 DE (CD93 antigen) (CDw93).
 GN C1QR1 OR CD93.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX PubMed=9047234;
 RA Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.;
 RT "CDNA cloning and primary structure analysis of C1qR(P), the human
 RT C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.";
 RL Immunity 6:119-129(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT ALA-318.
 RX PubMed=11781389;
 RA Steinberger P., Szekeres A., Wille S., Stockl J., Selenko N.,
 RA Prager E., Staffler G., Madic O., Stockinger H., Knapp W.;
 RT "Identification of human CD93 as the phagocytic c1q receptor (C1qR(p))
 RT by expression cloning.";
 RL J. Leukoc. Biol. 71:133-140(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead D.M., Beate D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownteen R., Sims S.,
 Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 RL [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leukocyte;
 RC Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RP CHARACTERIZATION.
 RP PubMed=11994479;
 RA McGreal E.P., Ikewaki N., Akatsu H., Morgan B.P., Gasque P.;
 RT "Human C1qR is identical with CD93 and the mN1-11 antigen but does
 RT not bind C1q.";
 RL J. Immunol. 168:5222-5232(2002).
 RN [6]
 RP O-GLYCOSYLATION.
 RX PubMed=10092817;
 RA Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.;
 RT "C1qR is a heavily O-glycosylated cell surface protein involved in
 RT the regulation of phagocytic activity.";
 RL J. Immunol. 162:3583-3589(1999).
 CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
 CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
 CC protein A (SPA). May mediate the enhancement of phagocytosis in
 CC monocytes and macrophages upon interaction with soluble defense
 CC collagens. May play a role in intercellular adhesion.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
 CC platelets, cells of myeloid origin, such as monocytes and
 CC neutrophils. Not expressed in cells of lymphoid origin.
 CC -!- PTM: N- and O-glycosylated.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -!- CAUTION: Has been sometimes referred to as a collectin receptor.
 CC -!- CAUTION: According to Ref.5, C1q is not a ligand for C1QR1.
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 3:1-6(2001);
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/467246456.g.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U94333; AAB53110.1;
 CC EMBL; AL118508; CAC00597.1;
 CC EMBL; BC028075; AAB28075.1;
 CC HSSP: P35555; 1EMN.
 CC MIM: 120377;
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000563; EGF-like.
 CC InterPro: IPR001881; EGF-Ca.
 CC InterPro: IPR001304; Lectin_C.
 CC InterPro: IPR001187; Tissue_factor.
 CC Pfam: PF00008; EGF; 5.
 CC Pfam: PF00059; Lectin_C; 1.
 CC Pfam: PF01108; Tissue_fac; 1.
 CC SMART: SM00034; CLECT; 1.
 CC SMART: SM00179; EGF_CA; 3.
 CC SMART: SM00001; EGF_like; 2.
 CC PROSITE: PS00610; ASX_HYDROXYL; 3.
 CC PROSITE: PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
 CC PROSITE: PS00641; C_TYPE_LECTIN_2; 1.
 CC PROSITE: PS01186; EGF_2; 3.


```
QY 845 ACAGGTGATGATGAGCTCTCCCATCAGCTGTCAATGGGGTGGCTTTCCTCTGGACC 904
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 135 lnArgLeuGlnAsnValGlnValAsnGlnLysValGlyMet---PheGluAlaHisI 154
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 905 TCGCTGCAAGAGGCGCACCCAAAATAACTCTGTCTGGGACTGTGACTGCTCTCCAGCGG 964
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 154 leGlnAlaGlnSer-----SerAlaIleGlnAlaProArgSerP 167
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 965 GCACITCTGCACAGCTGCAGCTGCAACACCTGCAGCTGCAGCTGCAGCTTCAAAGC 1024
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 167 roArgLeuGlyArgAlaArgSerProSerProCysProPheArgSerSer---SerGlnP 186
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1025 CATAAGGCGTCTCTGATAGAAATCCTGAAGCTTTCAACCAAAAAA-----TGGG 1075
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 186 roProGlyArgValLeuValGlnGlnAlaArgSerGluGluArgArgThrLysSerTrpG 206
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1076 GAAAGCGCTGCGGAGCTGCTAAACTTTCGACACAGCA----- 1113
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 206 lyGluGlnCysProGluThrSerGlyThrAspSerGlyArgLysGlyProSerLeuC 226
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1114 -----AAGGTGCACTGTAAAGCGCTCAGGCTGCCTGCAAGAACTACTG 1156
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 226 ysSerSerGlnValLysLysGlyMetProProLeuProGlyArgAlaAlaProThrGlyS 246
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1157 TGAGTGTCTATGAGCGCAAAATCATGTGT-----CTTCCATTGCAAAATG 1201
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 246 erGluAlaGlnGlyProSerAlaPheValArgMetGluLysGlyIleProAlaSerProA 266
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1202 CATTCCTGCAAAAATATGAAGAAGTCCAGAACCAAAATGCTGTAGCAGCACCCCA 1261
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 266 rgCysGlySerProThrAlaMetGluIleAspLysArg-----GlySerProT 282
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1262 CTACATGG-----AGCCTGGGACTTTGAGACAGCACTT----- 1296
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 282 hrProGlyThrArgSerCysLeuAlaProSerLeuGlyLeuPheGlyAlaSerLeuThrM 302
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1297 -----ATTGTCCCCAGCAAGTTCT-----CAGGACCTCC 1327
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 302 etAlaThrGluValAlaAlaArgValThrSerThrGlyProHisArgProGlnAspLeuA 322
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1328 AAACTGTGAAAAATATGACAG-----CCTCTCTCTGTAT 1363
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 322 laLeuThrGluProSerGlyArgAlaArgGluLeuGluAspLeuGlnProProGluAla- 341
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1364 CTCCTGGGAAGTAGTGA-----GGCCACATGTGCTGCTGGCCCGGAGGTGA 1414
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 342 LeuValGluArgGlnGlyGlnPheLeuGlySerGluThrSerProAlaProGluArgGly 361
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1415 GGAACGACAGCAGGACACTGTTCCTCCCAAGCTTGCTGAGCAGATGATCTGGAGGAGTT 1474
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 362 GlyProArgAspGlyGluProProGlyLysMetGlyLysGlyTyLeuProCysGly--- 380
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1475 TGAAGGTGCTGCTGCCAGATCTCCACAT-----CGAGTT 1510
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 381 -----MetProGlySerGlyGluProGluValGlyLysArgProGluGluThrVal 398
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1511 CAAGTCCAAAGGCGTGAATAATGACTAGCTGCAAGCTGGTAAAGGGAATGCTGTGGC 1570
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 399 serValGlnSerAlaGlnSerSerAspAla-LeuSerTrpSerArg---LeuProArgAl 417
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1571 AAGCCTCAGCCCTGGGAATCTGCACCGAAGAGTGTGCTCCCGAGGAGGAGCGCGC 1630
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 417 aLeuAlaSerValGly-----ProGluGluAlaArgSerG 429
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1631 CCATCATGCGGCAGCTCAGCTGTGAGTCTGAGTGTATGATGTGCTACTGCGCCACTAC 1690
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 429 yAlaProValGlyGlyGlyArgTrpGlnLeuSerAspArgValGluGlyGlySerProTh 449
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1691 TCAAGGTATCTCTA-----AAGTGAACGACGAGCAGCCCTGGGGATGACACTGGCCCT 1747
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 449 rLeuGlyLeuLeuGlyGlySerProSerAlaGlnProGlyThrGly-----As 465
```

```
QY 1748 CQTGTCCCTGGGAGGAGCCCTCTGGG-----GACTCCTGCTGCATAAAAGAGG 1798
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 465 nValGluAlaGlyIleProSerGlyArgMetLeuGluProLeuProCys----- 481
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1799 GTGATTTTCTACTTCTGTTATGTCTTTGCTTTCAAAATGCTTAGTACCTCCATTCA 1858
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 482 -----TrpAspAlaAl 485
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1859 AGTTATTATGAGCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTGCGCC 1918
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 485 aLysAspLeuGlyGluProGln-----CysPr 494
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1919 AAATCACATACAAGTCAGGTGCCATCAGGGTGTTCAGGCCAGGCTGTGCACAGGAG 1978
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 494 oProGlyAspArgVal-----GlyValGlnProGlyAsnSer-----Arg-- 507
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1979 ATATGGGAGGGGGTGGGTAGAGCTGGTGTGGTGTGGTGG 2017
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 508 ValTrpGlnGlyThrMetGluLysAlaGlyLeuAlaTrp 520
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
RESULT 15
OC90_HUMAN STANDARD; PRT; 493 AA.
ID OC90_HUMAN
AC Q02509;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Otoponin 90 precursor (Oc90) (Phospholipase A2 homolog).
GN OC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Murthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RL A2 related gene.";
RL Nucleic Acids Res. 21:135-143(1993).
CC -1- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z14310; CAA78662.1; ALT_INIT.
DR HSSP; P00593; 4BP2.
DR Genew; HGNC:8100; OC90.
DR MIM; 601658;
DR InterPro; IPR001211; PhospholipaseA2.
DR Pfam; PF00068; Phoslip; 2.
DR PRINTS; PR00389; PHPLIPASEA2.
DR PRODOM; PD000303; PhospholipaseA2; 2.
DR SMART; SM00085; PA2C; 2.
DR PROSITE; PS00118; PA2_HIS; 2.
DR PROSITE; PS00119; PA2_ASP; 1.
DR Signal; Glycoprotein; Repeat.
KW SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 493 OTOCONIN 90.
FT DOMAIN 76 190 PHOSPHOLIPASE A2-LIKE 1.
FT DOMAIN 321 377 PHOSPHOLIPASE A2-LIKE 2.
FT DOMAIN 389 441 PHOSPHOLIPASE A2-LIKE 3.
FT DISULFID 85 145 BY SIMILARITY.
FT DISULFID 99 190 BY SIMILARITY.
```

Db	341	SerTyrGlyCysTyrCysGlyGlnGluGlyArgGlyGluProArgAspLeuAspArg	360
		::::	
QY	1140	--TGCCTGAAGAACTACTGTGTAGTGCTATGAG-----	1169
		:::::	
Db	361	CysCysLeuSerHisHisCys---CysLeuGluGlnValArgArgLeuGlyCysLeuLeu	379
		:::::	
QY	1170	-----GCCAAATCATGTGT-----TCT	1187
		::: :::::	
Db	380	GluArgLeuProTrpSerProValValCysValAspHisThrProLysCysGlyGlyGln	399
		TCCATTTGCCAAATGCATGTCTGCAAAAACTATGAAGAAAGTCCAGAACGAAAAATGCTG	1247
		::: ::::	
Db	400	SerLeuCysGluLysLeuLeuCysAlaCysAspGlnThrAlaAlaGlu-----	415
		:::	
QY	1248	ATGAGCACACCCCACTACATGAGGAGCTGGGGACTTTGAGAGCAGCCATATTATTGTCGCCA	1307
Db	416	-----CysMetThrSerAlaSerPheAsnGlnSer-----	425
QY	1308	GCCAAAGTCTCAGGAGCTCCAAACCTGAGAAATAATAGCAGGCGCTTCTCCTGTATCTCC	1367
		:::: ::::	
Db	426	-----LeuLysSerProSerArgLeuGlyCysProGlyGlnProAlaAlaCysGluAsp	443
QY	1368	TGGGAAGTAGTGGAGGCCCATGTGCTGCTGCTGGCCCGCCAGGGTGAGGAACGACAGCAG	1427
		::: ::::	
Db	444	---SerLeuHisProValProAlaAlaProThrLeuGlySerSerSerGluGluAspSer	462
		GAGCAGCTGTTCCCAAGCTTGGCTGAGCAGATGATCCTGTGGAGGAGTTTGGAGGTGCTG	1487
		::::	
Db	463	GluGluAspProProGln-----GluAspLeuGlyArg-Ala--	474
QY	1488	TCGCAGATGTTCCACATCGAGTTTCAAGTCCAAAGGGC	1524
		:::	
Db	475	-LysArgPheLeuArgLysSerLeuGlyProLeuGly	486

Search completed: April 21, 2003, 11:29:51
Job time : 64.0475 secs

GenCore version 5.1.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:20:22 ; Search time 141.77 Seconds
(without alignments)
5514.106 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tatctctgtgggtggccgcg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p.model -DEV=xlp
-Q/cgn21/uspto.spool/US09743237/runat_21042003_111944_399/app_query.fasta_1.4494
-DB=SPTRMBL_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237@cgn.1.1.462@runat_21042003_111944_399 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rviro: *
16: sp_bacteriap: *
17: sp_archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	1881.5	46.9	475 11 Q8VIE0	Q8vie0 mus musculus

2	1871.5	46.7	475 11 Q8VIE1	Q8vie1 mus musculus
3	786.5	19.6	277 11 Q9D571	Q9d571 mus musculus
4	440.5	11.0	5 09V6Q8	Q9v6q8 drosophila
5	400.5	10.0	429 5 062295	Q62295 caenorhabdi
6	400	10.0	435 5 095QD7	Q95qd7 caenorhabdi
7	371	9.3	571 10 Q9SL70	Q9sl70 arabidopsis
8	366	9.1	603 10 Q9SZD1	Q9szd1 arabidopsis
9	266.5	6.6	601 10 Q9CAV1	Q9cav1 arabidopsis
10	253.5	6.3	356 10 Q9LW71	Q9lw71 arabidopsis
11	248.5	6.2	896 10 Q9ZS22	Q9zs22 glycine max
12	246.5	6.2	526 10 Q94AI2	Q94ai2 arabidopsis
13	228	5.7	695 10 Q9LE32	Q9le32 arabidopsis
14	228	5.7	695 10 Q9LUJ3	Q9luj3 arabidopsis
15	228	5.7	695 10 Q9M679	Q9m679 arabidopsis
16	226.5	5.7	609 10 Q9VMQ3	Q9vmq3 drosophila
17	186	4.6	243 5 Q9VUQ3	Q9vlu5 arabidopsis
18	183.5	4.6	593 10 Q9LU15	Q9lds2 oryza sativ
19	168.5	4.2	553 10 Q94DS2	P70200 mus musculu
20	150.5	3.7	1175 11 P70200	O35126 mus musculu
21	146.5	3.6	1175 11 O35126	O75097 homo sapien
22	146	3.6	1737 4 O75097	O44341 hallotis ru
23	131.5	3.3	1428 5 O44341	O75851 homo sapien
24	131.5	3.3	4123 4 O75851	Q9ul14 homo sapien
25	126.5	3.2	1481 4 Q9UL14	Q9u495 homo sapien
26	125.5	3.1	1182 4 Q9U495	Q9v08 drosophila
27	124	3.1	3680 5 Q9V08	Q95428 homo sapien
28	123.5	3.0	1235 4 Q95428	O14687 homo sapien
29	123.5	3.0	4957 4 O14687	O14686 homo sapien
30	123.5	3.0	5262 4 O14686	Q99621 homo sapien
31	123	3.0	1190 4 Q99621	Q9hlv5 homo sapien
32	122.5	3.1	1130 4 Q9HLV5	Q96dn2 homo sapien
33	121	3.0	955 4 Q96DN2	Q8spm4 bos taurus
34	120.5	3.0	5146 6 Q8SPM4	Q9m1x1 arabidopsis
35	119.5	3.0	975 10 Q9M1X1	Q9y6r7 homo sapien
36	119.5	3.0	2843 4 Q9Y6R7	Q951g1 equus cabal
37	118.5	3.0	610 6 Q95LG1	O88281 rattus norv
38	117.5	2.9	1574 11 O88281	Q9y5l9 homo sapien
39	117.5	2.9	2971 4 Q9Y5L9	Q9y075 leishmania
40	116	2.9	383 5 Q9Y075	Q9luw0 oryza sativ
41	116	2.9	544 10 Q9LUW0	Q9jl72 mus musculus
42	115.5	2.9	938 11 Q9JL72	Q9438 chironomus
43	115	2.8	1698 5 Q9438	Q93321 fugu rubrip
44	115	2.8	4823 13 Q93321	O57580 gallus gall
45	114.5	2.9	1151 13 O57580	

ALIGNMENTS

RESULT 1
Q8VIE0 PRELIMINARY; PRT; 475 AA.
AC Q8VIE0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tesmin (Hypothetical 50.6 kda protein).
GN TESM
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AB057423; BAB64935.1; -
DR EMBL; BC024377; AAH24377.1; -
DR InterPro; IPR005172; CXG.
DR Pfam; PF03638; CXG; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF79F CRC64;

Alignment Scores:
Pred. No.: 5,76e-173 Length: 475
Score: 1881.50 Matches: 353
Percent Similarity: 96.47% Conservative: 2
Best Local Similarity: 95.92% Mismatches: 6
Query Match: 46.94% Indels: 7
DB: 11 Gaps: 1

US-09-743-237-1 (1-2241) x Q8VIE0 (1-475)
QY 453 CTACTGCCAGGTCCAGGCGAGCAAGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAAATCAAAAGACAGCGTGTAGTGTGCCAGCGCGAGCGCTTAAGACGACGCT 551
Db 128 ProValGluIleLysGluAlaGlySerValProGlyGlySerProGlyAspAlaAla 147
QY 552 TTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTTCCCATCATCCAGGAGGAGAG 611
Db 148 PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCCCTCGGGAAGAAAGACTCCAGCCCATCGTGTATTTGTCAGTGA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerProMetValIleCysGlnLeuLys 187
QY 672 GGAGCGCCAGACTCTGTGATACAACTGTGCGCGAGGAGCTCAAGCGCTCCAT 731
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCTCAGTACGAGCAGCAGTTCCTCCTCAGTACAGCTCCCTAAGCAATG 791
Db 208 LeuLeuProGlnInTyAspAspGlnSerSerPheProGlnSerGluLeuProLysProMet 227
QY 792 ACACTTTAGTGGGGAAGACTTCTGCCAGTACCAGCAAGTAAATCTCATCACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGAGCTCTCCCATCAGCTCAATGGGGCTGCTTCCCTCTGGACCTCTCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGCCACCCAAATCTGTCTGGTACTGTGACTGCTTCCAGCGGGGACTTC 971
Db 268 GlnGlyProProLysIleThrLeuSerGlyTyrcysAspCysPheSerSerGlyAspPhe 287
QY 972 TCCACAGCTGAGCTGCAACACTCGCCCATGAGCTGAGCTGAGCGCTTCAAGCCATAAG 1031
Db 288 CysAsnSerCysSerCysAsnAsnLeuArgHisGluLeuLeuArgPheLysAlaIleLys 307
QY 1032 GGTGCTCTTGATAGAAATCTGAAGCTTCCAAACCAAAATGGGAAAGCGCTGGGA 1091
Db 308 AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly 327
QY 1092 GCTGCTAACTTCGACAGCAAGAGGTGCAACTGTAAGCGCTCAGCGTGTGCTGAAGAAC 1151
Db 328 AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn 347
QY 1152 TACTGTGAGTGTATGAGCCAAATCATGTCTTCCATTTGCAATGATGCTTGC 1211
Db 348 TyrCysGluCysTyrgluAlaLysIleMetCysSerSerIleCysLysCysIleAlaCys 367
QY 1212 AAAAATATGAAGATCCAGAACGAAATGCTGATGAGCACACACCCCACTACATGAG 1271
Db 368 LysAsnTyrgluGluSerProGluArgLysMetLeuMetSerThrProHisTyrgluMetGlu 387

RESULT 2
Q8VIE1 PRELIMINARY; PRT; 475 AA.
ID Q8VIE1
AC Q8VIE1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tesmin.
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein: a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057422; BAB64934.1; -
DR InterPro; IPR005172; CXG.
DR Pfam; PF03638; CXG; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Alignment Scores:
Pred. No.: 5,35e-172 Length: 475
Score: 1871.50 Matches: 351
Percent Similarity: 95.92% Conservative: 2
Best Local Similarity: 95.38% Mismatches: 8
Query Match: 46.69% Indels: 7
DB: 11 Gaps: 1

US-09-743-237-1 (1-2241) x Q8VIE1 (1-475)
QY 453 CTACTGCCAGGTCCCAAGGAGCAAGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAAATCAAAAGACAGGTTGTTAGTGTGCCAGCGCGAGCGCTTAAGACGACGCT 551
Db 128 ProValGluIleLysGluAlaGlySerValProGlyGlySerProGlyAspAlaAla 147
QY 552 TTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTTCCCATCATCCAGGAGGAGAG 611
Db 148 PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCCCTCGGGAAGAAAGACTCCAGCCCATCGTGTATTTGTCAGTGA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerProMetValIleCysGlnLeuLys 187
QY 672 GGAGCGCCAGACTCTGTGATACAACTGTGCGCGAGGAGCTCAAGCGCTCCAT 731
```



```
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCAGTACGATGACAGACAGAGTTCCTCCAGTACGAGCTCCCTAAAGCAATG 791
Db 208 LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysProMet 227
QY 792 ACAACCTTTAGTGGAGAGCTTCGCGAGTACACGAGGAGTAAATCTCATCACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGGAGCTCTCCATCAGCTGCAATGGGCGCTCCCTTCCCTCTGGACCTGCTCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGGCCACCAATAACTCTGCTGGTACTGTGACTGTCTTCCAGCGGGGACTTC 971
Db 268 GlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysPheSerGlyAspPhe 287
QY 972 TGCAACAGCTGACGCTGCAACACCTGCGCCATGAGCTCGAGCGCTTCAAGGCCATAAAG 1031
Db 288 CysAsnSerCysSerCysAsnAsnLeuArgHisGluLeuGluArgPheLysAlaIleLys 307
QY 1032 GCGTGTCTGTATGAAATCTGAAGCTTCCCAACCAAAATGGGAAAGCCGCTCTGGGA 1091
Db 308 AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly 327
QY 1092 GCTGCTAAATCTCGACACGCAAGGTCGAACCTGTAAAGCGCTCAGCGCTCGAGAAC 1151
Db 328 AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn 347
QY 1152 TACTGTGAGTGCATGAGGCAAAATCATGTGTTCTTCCATTTGCAAAATGCATGCTTGC 1211
Db 348 TyrCysGluCysTyrGluAlaLysIleMetCysSerSerIleCysLysCysIleAlaCys 367
QY 1212 AAAACTATGAAGAAGTCCAGAACGAAATAATGCTGATGAGCACACCCCACTACATGGAG 1271
Db 368 LysAsnTyrGluGluSerProGluArgLysMetLeuMetSerThrProHisTyrMetGlu 387
QY 1272 CTTGGGAGCTTTCAGAGCAGCAATTTCTCCCGCCAGCCAAAGTCTCAGGACCTCCAAA 1331
Db 388 ProGlyAspPheGluSerHisTyrLeuSerProAlaLysPheSerGlyProProLys 407
QY 1332 CTGAGAAAAATAGGAGGCTTCTCTGTATCTCTGGAAGTGTAGTGGAGGCGCACATGT 1391
Db 408 LeuArgLysAsnArgGlnAlaPheSerCysIleSerTrpGluValGluAlaThrCys 427
QY 1392 GCTGCTGCTGCTGCGCCAGGTGAGGAGGAGCAGAGCAGACACTGTTCCCAAGCTTGGCT 1451
Db 428 AlaCysLeuLeuAlaHisGlyGluGluAlaGluHisGluHisCysSerProSerLeuAla 447
QY 1452 GAGCAGATGATCTCGAGGAGTTTGAAGTGCTGCTCCAGATTCTCCACATCGAGTTC 1511
Db 448 GluGlnMetIleLeuGluGluPheGlyArgCysLeuSerGlnIleLeuHisIleGluPhe 467
QY 1512 AAGTCCAAGGGCTGAAATTTGAG 1535
Db 468 LysSerLysGlyLeuLysIleGlu 475
```

RESULT 3

```
Q9D571 ID Q9D571 PRELIMINARY; PRT; 277 AA.
AC Q9D571;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4930509C02, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1; -
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;
```

Alignment Scores:

```
Pred. No.: 4,52e-67 Length: 277
Score: 786.50 Matches: 155
Percent Similarity: 92.35% Conservative: 2
Best Local Similarity: 91.18% Mismatches: 6
Query Match: 19.62% Indels: 7
DB: 11 Gaps: 1
```

US-09-743-237-1 (1-2241) x Q9D571 (1-277)

```
QY 453 CTACTGCCAGTTCCCAAGGCGACGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAATAACAAGAGCAGGTGTGTGTCAGAGGCGGCGCTGAAAGCGAGCT 551
Db 128 ProValGluIleLysGluAlaGlyGlySerValProGlyGlySerProGluAspAlaAla 147
QY 552 TTCCAGGCGCCCTCGCTCAGGAATCCTGTGCAAGTTCCTCATCATCCAGGAGGCGAG 611
Db 148 PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCGCTCGGAAGAAGACTCCAGCGCCCATGCTGATTTCTCAGCTGAAA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerPrometValIleCysGlnLeuLys 187
QY 672 GGAGGCGCCAGATGCTCTGCATAGACACTGTGGCGGCGAGGAGCTCAAAGCGCTCCAT 731
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCAGTACGATGACGAGCAGTTTCCCTCAGTCAGAGCTCCCTTAAGCCAATG 791
Db 208 LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysPromet 227
QY 792 ACAACTTTAGTGGGAAGACTTCTGCGAGTACACGCAAGTAAATCTCATCACACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGGAGCTCTCCATCAGCTGTCATGAGTGGGCGCTTCCCTCTTGACCTGCTGCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGGCCACCAATAACTCTGCTGGG 941
Db 268 GlnGlyProProLysIleThrLeuSerGly 277
RESULT 4
```



```

QY 1350 -----GCTTCTCTCTGATCTCTGGAAGTAGTGAGGCGCACATGTGCTGCT 1399
      |||
      : : : : : |||
Db 878 nGluAsnArgAlaAsnIleTyrPheThrAspValIleGluAlaThrIleMetCysMe 898
      : : : : : |||
QY 1400 GCTGCGCCAGGTGAGGAAGACAGACAGACACTGTTCCTCCCAAGCTTGCTGAGCAGAT 1459
      : : : : : |||
      : : : : : |||
Db 898 tileSerGluIleValMetHisGluTysGlnAsnValAlaValGluAspMetGluArgG1 918
      : : : : : |||
QY 1460 GATCTGAGGAGGAGTTGGAAGTGTCTGTCGCGAGATCTC 1499
      : : : : : |||
      : : : : : |||
Db 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931

RESULT 5
O62295
ID O62295 PRELIMINARY; PRT; 429 AA.
AC O62295;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE JCB.6b protein.
GN JCB.6b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RT Science 282:2012-2018(1998).
DR EMBL; 282274; CAB05228.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

Alignment Scores:
Pred. No.: 1.17e-29 Length: 429
Score: 400.50 Matches: 131
Percent Similarity: 43.81% Conservative: 60
Best Local Similarity: 30.05% Mismatches: 136
Query Match: 9.99% Indels: 110
DB: 5 Gaps: 19

US-09-743-237-1 (1-2241) x O62295 (1-429)
QY 419 CAGGGGAAATCATTTTGGTGACATCGATGAACACTGCTCCAGGTTCCTCCCAAGGAGCAA 478
      |||
      : : : : : |||
      : : : : : |||
Db 3 GlnGlyGluIleVal-TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22
      : : : : : |||
QY 479 GCAAGCAAGAAAGTGTGTAATCAAGAACAGCGTGTACTGTG----- 524
      : : : : : |||
      : : : : : |||
Db 22 pAsnTyrGluGlyAlaGluPheIleGluValAsnGlyGlnLeuValProHisAsnPr 42
      : : : : : |||
QY 525 -----CCAGGCGGCGCCCTCGAAGACGAG----- 549
      : : : : : |||
      : : : : : |||
Db 42 oAlaGlnGlnAsnArgProGlyThrSerSer-MetIleGlnHisAsnArgSerMetG 62
      : : : : : |||
QY 550 ---CTTTCAGGCCCTCTGGCTCAGGAATCTGTTGCAAGTTCATCAT----- 597
      : : : : : |||
      : : : : : |||
Db 62 luValAsnGlnGlyLeuValAspGluProIleAspThrSerSerHisArgValTyrV 82
      : : : : : |||
QY 598 -----CCCAGGAGGACAGAGGCGCTCCAGCTGCCCTCGGAAGAAGACTCCAGCC 649
      : : : : : |||
      : : : : : |||
Db 82 alProProArgProValGlnArgLysPro-----GlyProSerThrPro----- 97
      : : : : : |||
QY 650 CATGCTGATTTCTCAGCTGAAGAGGCGGCCCGCCAGATGCTCTCATAGACAACTGTGGCGC 709

```

```

Db 97 ----- 97
QY 710 GAGGAGCTCAAGCGCTCCATCTGCTTCTCAGTACGATGACACAGACAGTTTCCTCA 769
      |||
      : : : : : |||
      : : : : : |||
Db 98 --GlySerGlnTyrThrValArgAsnLeuSerAsnLeuSerGlySerProSerMet 117
      : : : : : |||
      : : : : : |||
QY 770 G-----TCAGAGCTCCCTAAG-----CCAATGACAACCTTTAGTGGGAA 807
      : : : : : |||
      : : : : : |||
Db 117 yrAspArgGlnProAlaSerLeuProArgThrValGlnProMetGlyLeuGluMetGlyA 137
      : : : : : |||
      : : : : : |||
QY 808 GACTTCTGCCAGTACCAGCGAAGTTAAATCTCATCACACAGCTTGATAATGAGGCTCC 867
      : : : : : |||
      : : : : : |||
Db 137 snSerGluGlnArgLysValTyrIleAspMetLysAspHisValSerHisIleArgLeuL 157
      : : : : : |||
      : : : : : |||
QY 868 CATCAGCTGTCAATGGGCTGCTTCCCTCTGGACCTCTGCAAGGGCCACCCAAA 927
      : : : : : |||
      : : : : : |||
Db 157 ysThrLysLysValPheAla-----ProGlyGlnArgLysProCysAsnC 173
      : : : : : |||
      : : : : : |||
QY 928 TAACT-----CTGTCTGGGTACTGTGACTGTCTTCCAGCGGGGACTTCTGCA 975
      : : : : : |||
      : : : : : |||
Db 173 ysThrLysSerGlnCysLeuLysLeuTyrCysAspCysPheAlaAsnGlyGluPheCysA 193
      : : : : : |||
      : : : : : |||
QY 976 ACAGCTGCAGTCG-----AACAACTCGCCATCAGCTCGACGCTTCAAG 1023
      : : : : : |||
      : : : : : |||
Db 193 rgAspCysAsnCysLysAspCysHisAsnAsnIleGluTyrAspSerGlnArgSerLysA 213
      : : : : : |||
      : : : : : |||
QY 1024 CCATAAGCGCTGCTTGTAGTAAATCTCGAAGCTTTCACCAACAAAATGGGGAAGGCC 1083
      : : : : : |||
      : : : : : |||
Db 213 laIleArgGlnSerLeuGluArgAsnProAsnAlaPheLysProLysIleGlyIleAlaA 233
      : : : : : |||
      : : : : : |||
QY 1084 GTCTGGGAGCTCTAAACTTCGA-----CACAGCAAAAGGGTCAACTTAAGCGCTCAG 1137
      : : : : : |||
      : : : : : |||
Db 233 rgGlyGlyIleThrAspIleGluArgLeuHisGlnLysGlyCysLysLysSerG 253
      : : : : : |||
      : : : : : |||
QY 1138 GCTGCTGAAGAACTACTGTGAGTGTATGAGCCCAAAATCATGTCTTCCATTTGCA 1197
      : : : : : |||
      : : : : : |||
Db 253 lyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysValProCysThrAspArgCysL 273
      : : : : : |||
      : : : : : |||
QY 1198 AATGCTATGCTGCAAAAC-----TATGAAGAAAGT----- 1229
      : : : : : |||
      : : : : : |||
Db 273 ysCysLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSerGlyG 293
      : : : : : |||
      : : : : : |||
QY 1230 -----CCAGAACGAAAAATGCTGATGAGC-----ACAC 1257
      : : : : : |||
      : : : : : |||
Db 293 lyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAsnAlaSerThrAlaThrP 313
      : : : : : |||
      : : : : : |||
QY 1258 CCCACTACATGAGCGCTGGGAGCTTTTGAGAGCAGCCATTAT-----TTGTCCCCAG 1308
      : : : : : |||
      : : : : : |||
Db 313 roAsp---SerGlyProGlySerValValThrAspGluHisGlyAspAspTyrGluAspM 332
      : : : : : |||
      : : : : : |||
QY 1309 CCAAGTCTCAGGACCTCCAAACTGAGAAAAAATAGGAGGCGCTTCTCTATCTCTCT 1368
      : : : : : |||
      : : : : : |||
Db 332 etLeuLeuSerHisLysProLysValGluMetAspProArgArgPhe-----Prot 349
      : : : : : |||
      : : : : : |||
QY 1369 GG-----GAAGTAGTGGAGGCCACATGTCCCTGCTGCGCCAGGGTG 1413
      : : : : : |||
      : : : : : |||
Db 349 rPTyrTyrMetThrAspGluValValGluAlaAlaThrMetCysMetValAlaGlnAlaG 369
      : : : : : |||
      : : : : : |||
QY 1414 AGGAAGCA-----GAGCAGGAGCACTGTTCCTCCCAAGCTTGGCTGAGC 1455
      : : : : : |||
      : : : : : |||
Db 369 luGluAlaLeuAsnTyrGluLysValGlnThrGluAspGluLysLeuIleAsnMetGluL 389
      : : : : : |||
      : : : : : |||
QY 1456 AGATGATCTCGAGAGGATTTGGAAGTGTCTGCTCGCAGATTCTC 1499
      : : : : : |||
      : : : : : |||
Db 389 ysLeuValLeuArgGluPheGlyArgCysLeuGluGlnMetIle 403
      : : : : : |||
      : : : : : |||

RESULT 6
Q95QD7
ID Q95QD7 PRELIMINARY; PRT; 435 AA.
AC Q95QD7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

```



```
Db 45 AlaAlaProThrValValAlaThrSerValLysProIleLeu-----SerSerVal 62
Qy 765 CCTAGTCAGAGCTCCCTTAACCAATGACAACTTACTAGTGGGAAGACCTCTCCAGTACCA 824
Db 63 ProSerThrIleArgPro--GlyMetThrIleAlaIleGlyGlnValThrGlnValArg 81
Qy 825 GCGAAGTTAAATCTCATCACACAGGTTGATATGAGCTCTCCCATCAGCTGTCGAATGGG 884
Db 82 ProThrLeuProMetAlaThrThrMetSerAsnProProSerGlnSerGlnIleValAsn 101
Qy 885 GCT-----GCTTTCCTCTGGAGCTCTCTGCAAGGGCCA---CCCAATA 929
Db 102 AlaProIleArgHisProIleProGluSerProLysAlaArgGlyProArgProAsnVal 121
Qy 930 -----ACT 932
Db 122 GluGlyArgAspGlyThrProGlnLysLysGlnCysAsnCysLysHisSerArgCys 141
Qy 933 CTGTCTGGGTACTGTACTGCTTCTCCAGCGGGACTTCTGCAACAGCTGCAGCTGC--- 989
Db 142 LeuLysLeuTyrcysGluCysPheAlaSerGlyThrTyrcysAspGlyCysAsnCysVal 161
Qy 990 -----ACACACTCGCGCATGAGCTCGAGCGCTTCAAGCCATAAAGCGTGCTT 1040
Db 162 AsnCysPheAsnAsnValAspAsnGluProAlaArgArgGluAlaValGluAlaThrLeu 181
Qy 1041 GATAGAAATCTGAAGCTTTCACCAACCAAAATGGGAAA-----GGCGGT--- 1085
Db 182 GluArgAsnProPheAlaPheArgProLysIleAlaSerProHisGlyGlyArgAsp 201
Qy 1086 -----CTGGGAGCTGCTAAACTT-----CGACACAGCAAGGCTGCAACTGT 1127
Db 202 LysArgGluAspIleGlyGluValValLeuLeuGlyLysHisAsnLysGlyCysHisCys 221
Qy 1128 AAGCGCTCAGCGCTGCTGAAGAACTACTGTGAGTGCTATGAGGCAAAATCATGTGTTCT 1187
Db 222 LysLysSerGlyCysLeuLysTyrcysGluCysPheGlnAlaAsnIleLeuCysSer 241
Qy 1188 TCCATTTCGAATGCTGCTTGCAGAACTATGAAGAACTCCAGAACGAAAGAAATGCTG 1247
Db 242 GluAsnCysLysCysLeuAspCysLysAsnPheGluGlySerGluGluArgGlnAlaLeu 261
Qy 1248 ATGAGCACACCCAC-----TACATGGAG----- 1271
Db 262 PheHisGlyGluHisSerAsnHisMetAlaTyrlLeuGlnAlaAlaAsnAlaIle 281
Qy 1272 CTGGGGACTTGTAGACGAGCCATTATTG---TCCCGAGCCAGTCTTCAGGACCTCCA 1328
Db 282 ThrGlyAlaValGlySerSerGlyPheAlaProSerProAla-----Pro 296
Qy 1329 AAACGTGAGAAAAATAGGCAG 1349
Db 297 LysArgArgLysGlyGlnGlu 303
RESULT 9
Q9CAV1
ID Q9CAV1 PRELIMINARY; PRT; 601 AA.
AC Q9CAV1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 66.8 kDa protein.
GN T9114.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
```

```
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delsen M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Horischer K., Kauer G., Loehner T., H., Nordsiek G.,
RA Reicher J., Schaefer M., Schoen O., Bagues M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA De Haan M., Maarsee A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana";
RL Nature 408:820-822(2000).
DR EMBL; AC009465; AAG51411.1; -
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 66786 MW; 10848E970D81E022 CRC64;
```

```
Alignment Scores:
Pred. No.: 1,21e-16 Length: 601
Score: 266.50 Matches: 89
Percent Similarity: 37.77% Conservative: 33
Best Local Similarity: 27.55% Mismatches: 96
Query Match: 6.65% Indels: 105
DB: Gaps: 16
```

US-09-743-237-1 (1-2241) x Q9CAV1 (1-601)

```
Qy 495 GTTCAATCAAGAGCAGGTGGTGTGTCAGCGCGCAGCCCTGAAGACGAGCTTC 554
Db 236 ValGluGluArgGluGlySerCysSerVal-----GlnValAlaAla 249
Qy 555 CAGGCCCTCTGGCTCAGGAATCCTGT-----TGCAAGTTCCTCATCCAG 602
Db 250 GlyAlaProAspIleAsnLeuSerCysSerSerLysValAlaAlaIleAspSerThra 269
Qy 603 GAGCAGGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTATTGT 662
Db 270 GluAlaGluAspLysGluAsp-----LysAspLeuGlnPro----- 281
Qy 663 CAGCTGAAGAGGAGGCGCCAGACTCTCTGCATACACACTGTGCGCGGAGGAGCTCAA 722
Db 282 -----SerGlyLysGlnArgSerValArgArgArgCysLeuThrPheAspMetGly 298
Qy 723 GCGCTCCAT---CTGCTTCCTCAGTACGATGACCCAGAGAGCTTTCCTCCAG- 770
Db 299 GlySerHisLysArgIleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThr 318
Qy 771 -----TCAGAGCTCCCTAAGCCCAATGACAACTTTA-----GTGGGA 806
Db 319 SerIleAsnLysAlaProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAsp 338
Qy 807 AGACTTCTGCCAGTACCAGCGCAAGTTAAATCTCATCACAGGTTGATATGGAGCTCTC 866
Db 339 GluIleLeuProIleProArgThrIleGlyLeuHisLeu-----AsnGlyPheVal 355
```



```

[1]
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Stavoren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenecker T., Pohl T.M., Terryn N.,
RA Giehl J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Slivey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moeres T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Reckman S., Ansoorge W., Cooke R., Berger C.,
RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RA "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RL Nature 391:485-488(1998).
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z97337; CAB10256.1; -.
DR EMBL; AL161539; CAB78519.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
KW Hypothetical protein.
SQ SEQUENCE 658 AA; 72129 MW; 5BB99148E5E13A77 CRC64;

Alignment Scores:
Pred. No.: 6.84e-15 Length: 658
Score: 248.50 Matches: 107
Percent Similarity: 39.95% Conservative: 50
Best Local Similarity: 27.23% Mismatches: 160
Query Match: 6.20% Indels: 76
DB: 18 Gaps: 18

US-09-743-237-1 (1-2241) x 023333 (1-658)
QY 441 GACTCCGATGACTACTGCCA---GGTTCCCAAGCCAGCAAGCAAGCAAGAAAGTGTT 497
Db 186 AspAlaSerGluLeuLeuIlePheAspSerProAspAlaSerGluAlaPheArgCysPhe 205
QY 498 GAAATCAAGAGAGAGGTGGTAGTGTGCCAGCGCGACCTGAGAGCGAGCTTTCAG 557
Db 206 MetMetGlnArgAlaSerAsnSer-----GluAlaArgPheArg 218
QY 558 GCCCCTCTGGCTCAGGAATCTGTGTCAAGTTCCATCATCCAGGAGGAGGAGGCC 617
Db 219 AsnGlyValGluLysGlnThr---MetGlnHisAspSerAsnLysGluProGluSerAla 237
QY 618 TCCAGTCCCTCGCAAGAAAGACTCCAGCCCCCGTGGTGGATTTGTCAG-----665
Db 238 AsnAlaIleProTyrGluValAsnSerGly-----ValIleSerGlnAlaValSerLeu 255
QY 666 CTGAAGGAGGCGCCAGCATGCTCTGCATAGC-----698
Db 256 LeuHisArgGlyIleArgArgCysLeuAspPheGluMetProGlyAsnLysGlnThr 275
QY 699 -----ACTGTGGCGCGAGGAGCTCAAGCG-----725
Db 276 SerSerGluAsnAsnThrAlaAlaCysGluSerSerArgCysValValProSerIle 295
QY 726 ---CTCCATCTG-----CTTCCTCAGTACGATGACGAGCAGTTCCTCCAGTCA 773
Db 296 GlyLeuHisLeuAsnAlaIleLeuMetSerSerLysAspCysLysThrAsnValThrGln 315
QY 774 GAGCTCCCTAAGCCCAATGACAACATTTAGTGGGAGAGACTTCTGCCAGTACCAGCAAGTTA 833
Db 316 AspTyrSerCysSerAlaAsnIleGlnValGlyLeuGlnArgSerIleSerThrLeuGln 335
QY 834 AATCTCATCACACAGGTTGTAATAGGAGCTCTCCATCAGCTGTCATGGGGCTGCCTTT 893

```

```

Db 336 AspSerLeuAspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspVal-----353
QY 894 CCCTCTGGACCTGCTCTGCAA-----GGGCCACCAAAATAACTGTGCT 938
Db 354 ProValGluProAlaLeuGlnGluLeuAsnLeuSerSerProLysLys-----Lys 370
QY 939 GGGTACTGTGACTGCTTCTCCAGGGGAGCTTCTGC---AACAGCTGCAGCTGC-----989
Db 371 SerTyrCysGluCysPheAlaAlaGlyValTyrCysIleGluProCysSerCysIleAsp 390
QY 990 -----ACAACCTCGCGCATGAGCTCGAGCGCTTCAAGCCATAAAGCGCTGTTCAT 1043
Db 391 CysPheAsnLysProIleHisGluAspValValLeuAlaThrArgLysGlnIleGluSer 410
QY 1044 AGAATCTGAGCTTTCACCAACCAAAATGGGAAGCCCTGCTGGAGCT-----1094
Db 411 ArgAsnProLeuAlaPheAlaProLysValIleArgAsnSerAspSerValGlnGluThr 430
QY 1095 -----GCTAAACTTCGACACAGCAAGAGGTGCAACTGTAAAG 1130
Db 431 GlyAspAspAlaSerLysThrProAlaSerAlaArgHisLysArgGlyCysAsnCysLys 450
QY 1131 CGTCAAGCTGCTCGAAGAACTACTGTGAGCTGATGAGGCCAAATAATCATGTCTTCC 1190
Db 451 LysSerAsnCysLeuLysLysTyrCysGluCysTyrGlnGlyGlyValGlyCysSerIle 470
QY 1191 ATTTGCAAAATGCTTGTGCAAAACTATGAGAAAGTCCAGCAAGCAAAATGCTGATG 1250
Db 471 AsnCysArgCysGluGlyCysLysAsnAlaPheGlyArgLysAspGlySerSerIleAsp 490
QY 1251 AGCACACCCCTACTACATGGAGCTGGGAGCTTGTGAGAGCAGCATTAATTTGCCCCAGCC 1310
Db 491 MetGluAlaGluGlnGlu-----GluAsnGluThrSerGluLysSerArgThrAla 508
QY 1311 AAGTCTCAGGACCTCCAAAACACTGAGAAAATAAGGAGGAGCTTCT-----1356
Db 509 Lys-SerGln-----GlnAsnThrGluValLeuMetArgLysAspMetSerSerAlaLe 526
QY 1357 -CCTGTATCTCCTGGGAAGTAGTGGAGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1415
Db 526 uProThrThrProThrProIleTyrArgProGluLeuValGlnLeu---ProPheSerSe 545
QY 1416 GAAGCAGAGCAGGAGCAGCTGTTCGCCAGCTGGCTG 1452
Db 545 rSerLysAsnArgMetProProProGlnSerLeuLeu 557

RESULT 12
Q92S22
ID Q92S22 PRELIMINARY; PRT; 896 AA.
AC Q92S22;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine-rich polycomb-like protein.
GN CPPI.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanich C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,
RA Pihakski-Maunbach K., Marcker K.A., Jensen E.O.;
RT "CPPI, a novel type DNA-binding protein involved in the expression of
RL a soybean leghemoglobin c3 gene.";
DR EMBL; AJ010165; CAA09028.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

```



```

QY 918 -----CCACCCAAAATAACTGTGCTGGGTACTGTGACTGCTTCTCC 959
Db 231 CysLysArgCysAsnCysLysLysSerLysCysLeuLysLeuTyrCysGluCysPheAla 250
QY 960 AGCGGGGACTTCTGC---AACAGCTGCAGCTGC-----AACAACTCGCCCAT 1004
Db 251 AlaGlyValTyrCysIleGluProCysSerCysIleAspCysPheAsnLysProIleHis 270
QY 1005 GAGCTCGAGCCCTCAAGAGCCATAAAGCGCTGCTTGTATGAAATCTGAAGCTTTCCAA 1064
Db 271 GluGluThrValLeuAlaThrArgLysGlnIleGluSerArgAsnProLeuAlaPheAla 290
QY 1065 CCAAAATG-----GGGAAAGGCCGCTGGGA 1091
Db 291 ProLysValIleArgAsnAlaAspSerIleMetGluAlaSerAspAlaSerLysThr 310
QY 1092 GCTGCTAACTTCGACAGCAAGGTCACACTGTGAAGCGCTCAGGCTGCCTCAAGAAC 1151
Db 311 ProAlaSerAlaArgHisLysArgGlyCysAsnCysLysLysSerAsnCysMetLysLys 330
QY 1152 TACTGTGAGTCTATGAGGCCAAAATCATGTGTTCTTCCATTTGCAATGCCATTGCTGC 1211
Db 331 TyrCysGluCysTyrGlnGlyGlyValGlyCysSerMetAsnCysArgCysGluGlyCys 350
QY 1212 AAAAAC-----TATGAA 1223
Db 351 ThrAsnValPheGlyArgLysAspGlySerLeuLeuValIleMetGluSerLysLeuGlu 370
QY 1224 GAAAGTCCAGAA-----CGAAATACTGTATGACACACCCCTACATGAG- 1271
Db 371 GluAsnGlnGluThrTyrGluLysArgIleAlaLysIleGlnHisAsnValGluValSer 390
QY 1272 -----CTGGGGACTTTTCAGAGCAGCCATTATTGTCCTCCAGCCCAAG 1313
Db 391 LysGluValGluGlnAsnProSerSerAspGlnProSerThrProLeuProTyrArg 410
QY 1314 -----TTCTCAGGACCTCCAAACTCGAGAAAATAAGG 1346
Db 411 HisLeuValValHisGlnProPheLeuSerLysAsnArg 423

RESULT 14
Q9LE32 ID Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TSOL (Putative DNA binding protein).
GN TSOL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TSOL is a novel protein that modulates cytokinesis and cell expansion
RL in Arabidopsis."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song Y.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TSOL, an
RL Arabidopsis gene with cysteine-rich repeats."
DR EMBL; AF204059; AAF69124.1;
DR EMBL; AF206324; AAF27433.1;

```

```

DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Alignment Scores:
Pred. No.: 6,71e-13 Length: 695
Score: 228.00 Matches: 91
Percent Similarity: 35.58% Conservative: 46
Best Local Similarity: 23.64% Mismatches: 152
Query Match: 5.69% Indels: 96
DB: 10 Gaps: 14

US-09-743-237-1 (1-2241) x Q9LE32 (1-695)
QY 366 AACTTAGAGAACTCTGATTTCTCTCGCTCTACATCTCTGAGTCTAGGTCTACAGGGG 425
Db 246 AsnGluAsnGlnProLeuAlaValLeuPro----- 255
QY 426 AAATCATTTTGGTGAGACTCCGATGAATCTGCCAGGTTCCCAAGGCGCAGCAAGCA 485
Db 256 -----ThrAsnGluSerValPheAsnLeuHisArgGlyGlyMetArg 269
QY 486 AGAAAAAGTGTGAATCAAGAGAGCAGGTGGTAGTGTCCAGCGCGCAGCCCTGAAGAC 545
Db 270 ArgArgCysLeuAspPheGlu-----MetProGlyLysArgLysLysAsp 284
QY 546 GCAGCTTTTCAGGCGCCCTCTGGCTCAGGAATCTGTGTCAGAGTTCCCATCATCCAGGAG 605
Db 285 IleValAspAsp-----GlnGlnSerValCys-----AspAsnAsnValAla 298
QY 606 GCAGAGGAGGCGCTCCAGCTGC----- 626
Db 299 GlyGluSerSerSerCysValValProGlyIleGlyLeuHisLeuAsnAlaValAla 318
QY 627 CCTCGGAAGAAAGACTCCAGCCCATGGTGTGATTTGTAGCTGAAGAGGCGCCAGATG 686
Db 319 MetSerAlaLysAspSerAsnIleSerValIleHisGlyTyrSerIleSerGlyGluIle 338
QY 687 CTCTGCATAGACAACACTGTGGCGGAGGAGCTCAAAGCGCTCCATCTGCTTCTCAGTAC 746
Db 339 GlnLysSerPheSerGlySerThrThrProIleGlnSerGlnAspThrValGlnGluThr 358
QY 747 GATGACCAGAGC---AGTTTCCCTCAGTCTCAGAGCTCCCTTAAGCAACTGACAACTTTAGT 803
Db 359 SerAspGlnAlaGluAsnGluProValGluGluValProLysAlaLeuValPheProGlu 378
QY 804 GGAAGACTTCTGCCAGTACCAGCAAGTAAATCTCATCATCAGAGTTGATAATGAGAGCT 863
Db 379 LeuAsnLeuGlySerLeuLysLysLysMetArgLysSerGluGlnAlaGlyGluGlyGlu 398
QY 864 CTCCCATCAGCTGTCAATGGGGCTGCCTTTTCCCTCTGGACCTGCTCTGCAAGGCCACCC 923
Db 399 SerCysLysArgCysAsn-----CysLysLys 407
QY 924 AAATAACTCTGTGGGTACTGTGACTGCTTCTCCAGCGGGGACTTCTGC---AACAGC 980
Db 408 SerLysCysLeuLysLeuTyrCysGluCysPheAlaGlyValTyrCysIleGluPro 427
QY 981 TGCAGCTGC-----AACAACTCGCCATGAGCTCGAGCGCTCCAAAGCCATA 1028
Db 428 CysSerCysIleAspCysPheAsnLysProIleHisGluGluThrValLeuAlaThrArg 447
QY 1029 AAGCGCTGCTTGTATGAAATCTCTGAAGCTTTTCCAAACCAAAATG----- 1073
Db 448 LysGlnIleGluSerArgAsnProLeuAlaPheAlaProLysValIleArgAsnAlaAsp 467
QY 1074 -----GGGAAAGGCCGCTGGAGCTGCTAACTTCGACACAGCAAAA 1115
Db 468 SerIleMetGluAlaSerAspAspAlaSerLysThrProAlaSerAlaArgHisLysArg 487
QY 1116 GGGTGCACACTGTAAAGCGCTCAGGCTGCTGAAGAACTACTGTGAGTGTCTATGAGCCCAA 1175
Db 488 GlyCysAsnCysLysLysSerAsnCysMetLysLysTyrCysGluCysTyrGlnGlyGly 507

```

```

QY 1176 ATCATGTGTTCTTCCATTGCAATGCAATGCTTGCAGAAAC----- 1217
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 ValGlyCysSerMetAsnCysArgCysGluGlyCysThrAsnValPheGlyArgLysAsp 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1218 -----TATGAAGAAAGTCCAGAA-----CGAAAA 1241
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 GlySerLeuValIleMetGluSerLysLeuGluGluAsnGlnGluThrTyrGluLys 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1242 ATGCTGATGAGCACACCCCACTACATGGAG-----CCTGGG 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ArgIleAlaLysIleGlnHisValGluValSerLysGluValGluGlnAsnProSer 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1278 GACTTTGAGAGCAGCATATTATTGTCAGGCAAG-----TTCTCAGGACCTCCAAAA 1331
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 SerAspGlnProSerThrProLeuProTyrArgHisLeuValValHisGlnProPhe 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1332 CTGAGAAAAATAGG 1346
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 LeuSerLysAsnArg 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q9LUI3 PRELIMINARY; PRT; 695 AA.
AC Q9LUI3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DNA binding protein-like.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR EMBL; AB022223; BAB01253.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76276 MW; 34BBA0E450F6BCE1 CRC64;

Alignment Scores:
Pred. No.: 6,71e-13 Length: 695
Score: 228.00 Matches: 91
Percent Similarity: 35.58% Conservative: 46
Best Local Similarity: 23.64% Mismatches: 152
Query Match: 5.69% Indels: 96
DB: 10 Gaps: 14

US-09-743-237-1 (1-2241) x Q9LUI3 (1-695)
QY 366 ACTTAGGAGAACCTCTGATTCCTGCTCATCTACATCTGCTAGTGCTAGGAGGGG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 AsnGluAsnGlnProLeuAlaValLeuPro----- 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 AATCATTTTGTGTGAGACTCCCATGACTGCGAGGTCCAGGTTCCCAAGGCAAGCAAGCA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 -----ThrAsnGluSerValPheAsnLeuHisArgGlyGlyMetArg 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 AGAAAAAGTTGAATCAAGAAGCAGGTGTAGTGCCAGGCGGCAGCCTCGAAGAC 545
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: April 21, 2003, 11:38:24
Job time : 162.77 secs

```

Db 270 ArgArgCysLeuAspPheGlu-----MetProGlyLysArgLysLysAsp 284
QY 546 GCAGCTTTCCAGCGCCCTCTGCTCAGGAATCCTGTTGCAAGTCCATCATCCAGGAG 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 IleValAspAsp-----GlnGlnSerValCys-----AspAsnAsnValAla 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 GCAGAGGAGCGCTCCAGCTGC----- 626
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 GlyGluSerSerSerCysValValProGlyIleGlyLeuHisLeuAsnAlaValAla 318
QY 627 COTCGAAGAAAGACTCCAGCCCCATGTTGTTGCTAGCTGAAAGAGGCGCCAGATG 686
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 MetSerAlaLysAspSerAsnIleSerValIleHisGlyTyrSerIleSerGlyGluIle 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 687 CTCTCATAGACAACCTGTGGCGGAGCTCAAGCGCTCCATCTGCTCTCTAGTAC 746
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 GlnLysSerPheSerGlySerThrThrProIleGlnSerGlnAspThrValGlnGluThr 358
QY 747 GATGACCAAGAGC---AGTTTCCCTCAGTCAGAGCTCCCTAAGCCCAATGACAACCTTTAGTG 803
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 SerAspGlnAlaGluAsnGluProValGluGluValProLysAlaLeuValPheProGlu 378
QY 804 GGAAGACTTCTCCAGTACCAGCAAGTAAATCTCATCACACAGCTTGATATGGAGCT 863
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 LeuAsnLeuGlySerLeuLysLysMetArgLysSerGluGlnAlaGlyGluGlyGlu 398
QY 864 CTCCTCATCAGCTGTCATGGGGCTGCTTCCCTCTGGACCTGCTCTGCAAGGGCCACCC 923
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 SerCysLysArgCysAsn-----CysLysLys 407
QY 924 AAAATAACTCTCTGCTGGTACTGTGACTGCTTCCAGCGGGGACTTCTGC---AACAGC 980
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 SerLysCysLeuLysLeuTyrCysGluCysPheAlaIleGlyValTyrCysIleGluPro 427
QY 981 TCAGCTGC-----AACAACTGCGCCATGAGCTCGAGCGCTTCAAGGCCATA 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 CysSerCysIleAspCysPheAsnLysProIleHisGluGluThrValLeuAlaThrArg 447
QY 1029 AAGGGTGTCTTGATAGAAATCCTGAAGCTTTCCCAACCAAAAATG----- 1073
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 LysGlnIleGluSerArgAsnProLeuAlaPheAlaProLysValIleArgAsnAlaAsp 467
QY 1074 -----GGGAAAGGCCCTCTGGGAGCTGCTAAACTTCGACACAGCAAA 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 SerIleMetGluAlaSerAspAlaSerLysThrProAlaSerAlaArgHisLysArg 487
QY 1116 GGTCGCAACTGAAGCGCTCAGGCTGCCTGAAGAACTACTGTGAGTGTATGAGGCCAAA 1175
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 GlyCysAsnCysLysLysSerAsnCysMetLysLysTyrCysGluCysTyrGlnGlyGly 507
QY 1176 ATCATGTGTTCTTCCATTGCAATGCAATGCTTGTGCAAAAAC----- 1217
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 ValGlyCysSerMetAsnCysArgCysGluGlyCysThrAsnValPheGlyArgLysAsp 527
QY 1218 -----TATGAAGAAAGTCCAGAA-----CGAAAA 1241
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 GlySerLeuValIleMetGluSerLysLeuGluGluAsnGlnGluThrTyrGluLys 547
QY 1242 ATGCTGATGAGCACACCCCACTACATGGAG-----CCTGGG 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ArgIleAlaLysIleGlnHisAsnValGluValSerLysGluValGluGlnAsnProSer 567
QY 1278 GACTTTGAGAGCAGCATATTATTGTCAGGCAAG-----TTCTCAGGACCTCCAAAA 1331
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 SerAspGlnProSerThrProLeuProTyrArgHisLeuValValHisGlnProPhe 587
QY 1332 CTGAGAAAAATAGG 1346
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 LeuSerLysAsnArg 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:49:14 : Search time 92.5 Seconds
(without alignments)
6148.256 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 3824
Sequence: 1 aattcgggggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09743237/runat_21042003_112208_1454/app_query.fasta_1.2311
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=ptc -NORN=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237.ecgn_1_153 -runat_21042003_112208_1454 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	1647	43.1	299	21	AA168464	Human testis speci
2	1251	32.7	295	21	AA168463	Mouse testis speci
3	836.5	21.9	251	22	AA193348	Human polypeptide,
4	730	19.1	147	22	AB196025	Human testicular a
5	730	19.1	147	22	AA195330	Human reproductive
6	483.5	12.6	950	22	AB162035	Drosophila melanog
7	441	11.5	438	20	AB163392	Caenorhabditis ele
8	378.5	9.9	280	22	AA17958	Novel human diagno
9	278	7.3	53	22	AA176158	Human colon cancer
10	200	5.2	243	22	AB168888	Drosophila melanog
11	149	3.9	4561	22	AB130203	Novel human diagno
12	149	3.9	9222	22	AB121064	Novel human diagno
13	130.5	3.4	2743	23	AB181598	Human laminin alph
14	130.5	3.4	3695	23	AB181588	Human laminin alph
15	130.5	3.4	3696	23	AA17310	Human laminin alph
16	130.5	3.4	3705	23	AA17309	Human laminin alph
17	125	3.3	1037	22	AA17142	Mouse CRIM1 protei
18	124	3.2	3011	13	AA121519	Compiled HCV sequ
19	121	3.2	507	16	AA166631	HCV J1 NS3-NS4 dom
20	121	3.2	1017	22	AA159813	TuT protein #4.
21	121	3.2	1615	22	AA159826	Protein #3 encoded
22	120	3.1	3010	14	AA130616	Polypeptide coded
23	120	3.1	3011	14	AA131621	Hepatitis C virus
24	119.5	3.1	1035	22	AB166062	Drosophila melanog
25	119.5	3.1	3010	16	AA182694	Partial HCV non-st
26	119.5	3.1	3010	16	AA168864	Hepatitis C virus
27	119	3.1	1036	23	AA118852	Human pharmaceutical
28	118.5	3.1	1801	19	AA150895	Rat laminin B2 cha
29	118	3.1	732	22	AA161140	Human NOV10 protei
30	118	3.1	2515	22	AB171354	Drosophila melanog
31	117	3.1	1021	18	AA134481	HCV antigen combin
32	117	3.1	1021	19	AA140039	Fusion protein c20
33	117	3.1	1021	23	AA122050	PSOD/c200/core exp
34	117	3.1	1771	22	AB162631	HCV NS35 polypepti
35	117	3.1	1771	22	AA162632	HCV NS35 polypepti
36	117	3.1	1771	22	AA162634	Amino acid sequenc
37	117	3.1	1771	22	AA162635	Amino acid sequenc
38	117	3.1	1892	22	AA162636	Amino acid sequenc
39	117	3.1	1911	22	AA162638	Amino acid sequenc
40	117	3.1	1921	22	AA162639	Amino acid sequenc
41	117	3.1	1944	22	AA162637	Peptide encoded by
42	117	3.1	2261	10	AA190164	Sequence encoded i
43	117	3.1	2301	10	AA192047	HCV polypeptide 1.
44	117	3.1	2435	13	AA125135	Sequence encoded i
45	117	3.1	2436	10	AA192050	Sequence encoded i

ALIGNMENTS

RESULT 1
AA168464
ID AA168464 standard; Protein; 299 AA.
XX AA168464;
AC AA168464;
XX
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.
 XX XX 17-JUL-1998; 98JP-0219856.
 PR
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 XX Sugiuhara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX WPI: 2000-147785/13.
 DR N-PSDB; AA288156.
 XX
 PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 XX
 XX Claim 1; Page 50-52; 63pp; Japanese.
 XX
 CC The present sequence represents a male germ cell regulatory factor
 expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 XX Sequence 299 AA;
 SQ
 Alignment Scores:
 Pred. No.: 6.48e-151 Length: 299
 Score: 1647.00 Matches: 299
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.07% Indels: 0
 DB: 21 Gaps: 0

US-09-743-237-3 (1-2134) x AAY68464 (1-299)

Qy 407 ATGTGATATGCCAATTAAGAGGGGGCACAAATGCTATGTATAGACAATTTAGAACGA 466
 Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
 Qy 467 AGAAGCACTAAAGCAGCTCCATTGGTCTCCTCAGTATCAAGATCAAAATAATTATCTCAG 526
 Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
 Qy 527 TCAGATGTCCTAAACCAATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAAA 586
 Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
 Qy 587 TTAATCTCATTAACAACAACCTTGAGGAGCCCTTACATCGGTAGTCAACGGGTCTGCT 646
 Db 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
 Qy 647 TTCCCTCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGTACTGTGAC 706
 Db 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
 Qy 707 TGCTTTGCCAGTGGGACTTTTGCACAACTGCAATTAATTAATTTGTGCAACAACCTTG 766
 Db 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnAsnCysCysAsnLeu 120
 Qy 767 CATCATGATATTGAACGGTTTAAGCCATTAAAGCATGCTTTGTTAGAAATCCAGAGCT 826
 Db 121 HisHisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140
 Qy 827 TTCAGCCAAAATAATGGAGGGCCCAATGGGCAATGTCAGCCCTCAGCCACCAACAAAGG 886
 Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
 Qy 887 TGCAACTGCAGGAGTCCAGTCTGCTGAAGAAATTAACGAGTCTATGAGGCCCAAT 946
 Db 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
 Qy 947 ATGTGTTCTTCTATTTCGAATGCAATGCTGGTTGCCAAAATAATTATGAAGAAAGCCCAACGA 1006
 DB: 21 Gaps: 0

Db 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
 Qy 1007 AAGACACTAATGAGCATGCCAACTACATCAGACTGAGCTTTGGAGGAGCAGCAATTAC 1066
 Db 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
 Qy 1067 CTGCCACCAACGAAATTTTCAGGACTTCCAGGATTCACGATAGCGCGCTTCCTCA 1126
 Db 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
 Qy 1127 TGATCTCTCGGAGGTGGAGGCCACATGCCCTGCTGCTCAGGGAGAGAGAG 1186
 Db 241 CysIleSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
 Qy 1187 GCCGAGAAAGAACACTCTCCAACTGCTGCAGCAGCAGATGATCTCGGAGGAATTTGGA 1246
 Db 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGly 280
 Qy 1247 AGGTGCTTATCACAGATTCCTCCACACTGAGTTTAAATTAAGGATTTGAAAATGGAG 1303
 Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 2
 AAY68463
 ID AAY68463 standard; Protein: 295 AA.
 XX AC
 XX AAY68463;
 XX AC
 XX 25-APR-2000 (first entry)
 XX DE
 XX Mouse testis specific factor tesmin SEQ ID NO:4.
 XX KW
 XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 XX KW differentiation regulatory factor; male germ cell regulatory actor;
 XX KW germ cell differentiation; sterility.
 XX OS
 XX Mus musculus.
 XX PN
 XX WO200004147-A1.
 XX PD
 XX 27-JAN-2000.
 XX PF
 XX 16-JUL-1999; 99WO-JP03859.
 XX PR
 XX 17-JUL-1998; 98JP-0219856.
 XX PA
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX PI
 XX Sugiuhara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX DR
 XX WPI: 2000-147785/13.
 XX DR N-PSDB; AA288155, AA288157.
 XX PT
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 useful for investigation of germ cell differentiation and sterility -
 XX PS
 XX Claim 1; Page 47-49; 63pp; Japanese.
 XX CC
 XX The present sequence represents a male germ cell regulatory factor
 expressed specifically in spermatocytes, designated tesmin.
 XX CC can be used in the investigation of the mechanisms of germ cell
 XX CC differentiation and sterility.
 XX SQ
 XX Sequence 295 AA;
 Alignment Scores:
 Pred. No.: 1.92e-112 Length: 295
 Score: 1251.00 Matches: 228
 Percent Similarity: 84.95% Conservatives: 26
 Best Local Similarity: 76.25% Mismatches: 41
 Query Match: 32.71% Indels: 4
 DB: 21 Gaps: 1

US-09-743-237-3 (1-2134) x AAY68463 (1-295)

```
QY 407 ATGGTCATATGCCAATTTGAAGGGGCCACACAAATCTATGTATACAAATTTCTAGACA 466
Db 1 MetValIleCysGlnLeuIleGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTCTCTCAGTATCAAGATCAAAATTAATATCTACAG 526
Db 21 ArgGluLeuIleAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTAGTAGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerGluLeuProIlePheMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 587 TTAATCTCATATACACAACTTGGAGGAGCCCTTACCATCGGTAGTCAACGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 647 TTCCCTCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlyProAlaLeuGlnGlyProProIleThrLeuSerGlyTyrCysAsp 100
QY 707 TGTCTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTTGTTCAACAACCTG 766
Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSer-----CysAsnAsnLeu 116
QY 767 CATCATGATATCAACGGTTTAAGCCATTAAGGCATGCTTGGTAGAATCCAGAGCT 826
Db 117 ArgHisGluLeuGluArgPheIleLysAlaIleLysAlaCysLeuAspArgAsnProGluAla 136
QY 827 TTCACGCCAAAATTTGGGAGGCCCAATTTGGCAATGTCAAGCCCGACACACAAAGGG 886
Db 137 PheGlnProIleMetGlyGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGly 156
QY 887 TCGAATCGAGGAGGTGAGCTGCTCGAAGAAATTTACTGCGAGTGTATGAGGCCCAAT 946
Db 157 CysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIle 176
QY 947 ATGTGTTCTTATTTGCAAACTGATTTGTCGCAAAATTTATGCAAAAGCCAGACGA 1006
Db 177 MetCysSerSerIleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArg 196
QY 1007 AGACACTAATGAGCTGCAACATCATGACACTGGAGGTTTGAAGGCGAGCCATTAC 1066
Db 197 LysMetLeuMetSerThrProHisTyrMetGluProGlyAspPheGluSerHisTyr 216
QY 1067 CTGCCACCAAGCAATTTTCAGGACTTCCAAAGATTTCAGTCACATAGGCGGCTTCCCTCA 1126
Db 217 LeuSerProAlaLysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSer 236
QY 1127 TGCATCTCTGGGAGTGTGGAGGCCACATGCGCTGCTGCTGCTCAGGGGAGAG 1186
Db 237 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 256
QY 1187 GCGGAGAAACACACTGCTCCAAAGTGCCTGGCAGACAGATGATCTCGGAGGAATTTGA 1246
Db 257 AlaGluGlnGluHisCysSerProSerLeuAlaGluGlnMetIleLeuGluPheGly 276
QY 1247 AGTGCTTTATCACAGATTCTCCACTGAGTTAAATCTTAAGGGATTGAATGGAG 1303
Db 277 ArgCysLeuSerGlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
```

RESULT 3

AAM93348

ID AAM93348 standard; Protein; 251 AA.

XX AC

XX AAM93348;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide, SEQ ID NO: 2895.

XX

```
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94268.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;
XX
Alignment Scores:
Pred. No.: 3,36e-72 Length: 251
Score: 836.50 Matches: 167
Percent Similarity: 95.43% Conservative: 0
Best Local Similarity: 95.43% Mismatches: 5
Query Match: 21.88% Indels: 3
DB: 22 Gaps: 1
US-09-743-237-3 (1-2134) x AAM93348 (1-251)
QY 9 GTCAAGCGAAGCTCCGGGGGCGGACGAGCGGGGAGGCTCTCGGGGAGTACCCC 68
Db 79 ValIleAlaLysLeuAlaGlyGlyAspSerAspGlyGlyLeuLeuGlyIlePro 98
QY 69 GGATCCCGAGAGCTCAGCGGCTGGAGGAGCTCGCGTCTCGAGGCCCGCGAGCGCCC 128
Db 99 GlyIleProGluLeuSerAlaLeuGluAspValAlaLeuLeuGlnAlaProGlnPro 118.
QY 129 GCTGCAAGCTGCACTTCTCTGCTCTCTGCTGCTACCGCGCAGCCGAGCCCGGGTGT 188
Db 119 AlaCysAsnValHisPheLeuSerSerLeuLeuProAlaHisArgSerProAlaValLeu 138
QY 189 TGGCCCTTGGGGCGGCTGGTCTTGGCGAAGGAGGCTCCACCCCGGGCGTCCGCATGCC 248
Db 139 ---ProLeuGlyAlaTrpValLeu-GluGlyAlaSerHisProGlyValArgMetIlePr 157
QY 249 AGTTGAAATCAAGTGAAGTGGTACTACTACAGTAAATAATCCGGAAGAAAGCAACTT 308
Db 157 oValGluIleLysGlu-AlaGlyGlyThrThrThrSerAsnAsnProGluGluAlaThrL 177
QY 309 TGCAGAAATCTTCTTGCTCAGGAATCTCTGTTTSCAAGTTCCCATGTCGCCAGAACTAGAGG 368
```

```
Db 177 euGlnAsnLeuLeuAlaGlnGluSerCysLysPheProSerSerGlnGluLeuGluA 197
Oy 369 ATGCCTCTCTGCTTCTTAAGAAAGATTCCAAACCCCAATGGTGATATGCCAATTGAAG 428
Db 197 spAlaSerCysSerLeuLeuLysLysAspSerAsnPrometValIleCysGlnLeuLysG 217
Oy 429 GGGGCACAAATGCTATGTATAGACAATTCTAGAACAGAGAACTAAAGCACTCCCAAT 488
Db 217 lyGlyThrGlnMetLeuLysLysAspAsnSerArgThrArgGluLeuLysAlaLeuHisL 237
Oy 489 TGGTCTCTAGTATCAAGATCAAAATAATATCTACAGTCA 529
Db 237 euValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSer 250

RESULT 4
ABB96025
ID ABB96025 standard; Protein: 147 AA.
XX AC ABB96025;
XX XX
XX DT 21-JUN-2002 (first entry)
XX XX
XX DE Human testicular antigen SEQ ID NO: 1409.
XX KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic.
XX OS Homo sapiens.
XX XX
XX PN WO200155317-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01329.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231988.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
```



```
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483232/52.
XX
PT Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer
XX
PS ClalM 11; SEQ ID NO 1409; 766pp; English.
XX
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a protein of the
CC invention.
XX
SQ Sequence 147 AA;
Alignment Scores:
Pred. No.: 5,938-62 Length: 147
Score: 730.00 Matches: 144
Percent Similarity: 97.30% Conservative: 0
Best Local Similarity: 97.30% Mismatches: 3
Query Match: 19.09% Indels: 1
DB: Gaps: 0
US-09-743-237-3 (1-2134) x ABB96025 (1-147)
QY 229 CCGGGCGTCCGATGATCCAGTTGAAATCAAGGTAAGCAGGTGGTACTACTACAGTAA 288
Db 1 ProGlyValArgMetIleProValGluIleLysGlu-AlaGlyGlyThrThrSerAs 20
QY 289 TAATCGGAAGAACCACTTTCAGAAATCTTCTGCTCAGGAATCCTGTGCAAGTCCC 348
Db 20 nAsnProGluGluAlaThrLeuGlnAsnLeuLeuAlaGlnGluSerCysCysLysPhePr 40
QY 349 ATGGTCCCAAGCACTAGAGATGCTCTGCTGCTCTCTTAGAAGATCCCAACCAAT 408
Db 40 oSerSerGlnGluLeuGluAspAlaSerCysSerLeuLysLysAspSerAsnProMe 60
QY 409 GGTGATATCCCAATTGAAAGGGGGCACACAAATGCTATGTATAGACAATCTAGACAAG 468
```

```
Db 60 tValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThrAr 80
QY 469 AGAACTAAAGCACTCCATTGGTTCTCCTCAGTATCAAGATCAAAATAATATCTACAGTC 528
Db 80 gGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSe 100
QY 529 AGATGTCCCTAAACCAATGACTGCTTTAGTAGGAGATTTTCCCGAGCATCAACAAAT 588
Db 100 rAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLysLe 120
QY 589 AAATCTCATTACACAACTTGGAGGAGCCTTACCATCGTAGTCAACGGGTCTGCTTT 648
Db 120 uAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAlaPh 140
QY 649 CCCCTCGGATCAACTCTTCCA 670
Db 140 eProSerGlySerThr***Pro 147
RESULT 5
AAM95330
ID AAM95330 standard; Protein; 147 AA.
XX
AC AAM95330;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 3988.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
```

```
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-02311242.  
PR 08-SEP-2000; 2000US-02311243.  
PR 08-SEP-2000; 2000US-02311244.  
PR 08-SEP-2000; 2000US-02311244.  
PR 08-SEP-2000; 2000US-02311413.  
PR 08-SEP-2000; 2000US-02311414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-02311968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.
```

```
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.
```

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL01300.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition

Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention.

Sequence 147 AA;

Alignment Scores:
Pred. No.: 5.93e-62 Length: 147
Score: 730.00 Matches: 144
Percent Similarity: 97.30% Conservative: 0
Best Local Similarity: 97.30% Mismatches: 3
Query Match: 19.09% Indels: 1
DB: 22 Gaps: 0

US-09-743-237-3 (1-2134) x AAM95330 (1-147)

```
QY 229 CCGGGCTCGCATGATCCAGTTGAATCAAGTAGCAGGTGGTACTACTACAAGTAA 288  
Db 1 ProglyValArgMetIleProValGluIleLysGlu-AlaGlyGlyThrThrSerAs 20  
QY 289 TAATCCGGAAGAACACTTTGCAGATCTTCTGCTCAGAAATCCTGTGCAAGTTCCC 348  
Db 20 nAsnProGluGluAlaThrLeuGlnAsnLeuLeuAlaGlnGluSerCysCysPhePr 40  
QY 349 ATGTCTCCAGGAACACTAGAGGATGCTCTGCTCTTCTTCTTAAAGAAATTCACCAAT 408  
Db 40 oSerSerGlnGluLeuGluAspAlaSerCysSerLeuLysLysAspSerAsnProMe 60  
QY 409 GGTGATATGCCAATTTGAAGGGGGGCACAAATGCTATGTATAGACAATTTCTAGACAAG 468
```


Db 668 nLeuGlnAlaGlnAlaLysGlnArgIleArgGlnGlnLeuProThrGluGlnSerTh 688
 Qy 611 -----GAGGAGCCTTACCATCGGTA-----GTCAACGGGTC 642
 Db 688 rProIleLysValGluProLysLeuProThrLeuProProGlyValLysAlaAsnValPr 708
 Qy 643 TGCTTTCCCTCGGATCACTCTCCAGGACCAACCAATAACTTTGGCTGGG----- 697
 Db 708 oAlaLysProLeuPheGluValLeuLysProProAlaThrAlaAlaAlaGlyAlaVa 728
 Qy 697 ----- 697
 Db 728 lasProLeuGlyGlyMetThrSerArgArgLysHisCysAsnCysSerLysSerGlnCy 748
 Qy 698 -----TACTGTGACTCTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTA 747
 Db 748 sLeuLysLeuTyrcysAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLy 768
 Qy 748 TAATTGTTGCAACAACCTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGTCT 807
 Db 768 sAspCysPheAsnAsnLeuAspTyrcysGluValGluArgGluArgAlaIleArgSerCysLe 788
 Qy 808 TGGTAGAATCCAGAGCTTCCAGCCAAATAATTGGGAAGGCCCAATTGGCAATGTCAA 867
 Db 788 uAspArgAsnProSerAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetAr 808
 Qy 868 GCCCAGCACAAAGGGTCAACTGCAGGAGTGCAGGCTGCCTGAAGATTAATCTGCGA 927
 Db 808 g---LeuHisAsnLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrcysGl 827
 Qy 928 GTGCTATGAGCCCAATATTGTCTCTCTTCTTATTTGCAAAATGCATTTGGTTCACAAATTA 987
 Db 827 uCysTyrcysGluAlaLysIleProCysSerSerIleCysLysCysValGlyCysArgAsnMe 847
 Qy 988 TGAGAAGCCCAACAGACACATATGAGCATGCAACATGCAACATGACAGCTGGAGG 1047
 Db 847 tGluAspArgProAsp-----ValAspMetAspSerLeuAspGlyLeuMetGl 863
 Qy 1048 TTTGAGAGCGACCATTTACCTGCACCAACAAATTTTTCAGGACTTCCAGATTCAGTCA 1107
 Db 863 yValGluGlyGln-----LysLysAspLysAlaLysAsnLysGlnLeuAs 878
 Qy 1108 CGATAGGCGGCTTCCTCATCATCTCTCTGGAGGTGGTGGAGCCACATCGCCTGCCT 1167
 Db 878 nGluAsnArgAlaAsnIleTyrcysThrAspValIleGluAlaThrIleMetCysMe 898
 Qy 1168 CTTGCTCAGGAGGAGGCGGAGAGACACTGCTCCAAAGTGCCTGGCAGACAGAT 1227
 Db 898 tileSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGl 918
 Qy 1228 GATCCTGGAGGAATTTGAGAGTGCTTATCATCAGATTC 1267
 Db 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931
 RESULT 7
 AAW83392
 ID AAW83392 standard; Protein; 438 AA.
 XX AC AAW83392;
 XX AC AAW83392;
 DT 29-MAR-1999 (first entry)
 XX DE Caenorhabditis elegans synMuv protein LIN-54.
 XX KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;
 KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 KW cell proliferation; gene therapy.
 XX OS Caenorhabditis elegans.
 XX PN W09854299-A1.
 XX PD 03-DEC-1998.

XX 28-MAY-1998; 98WO-US11043.
 XX 28-MAY-1997; 97US-0047996.
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA Ceol C, Horvitz HR, Lu X;
 PI WPI; 1999-045362/04.
 XX DR N-PSDB; AAV72865.
 XX Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas
 XX Claim 7; Fig 13; 70pp; English.
 XX This is the amino acid sequence of LIN-54, a novel protein of
 CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
 CC proliferation, and is part of a pathway that may be used as a
 CC genetic and biochemical model system for tumour suppression and
 CC cancer in mammals. SynMuv pathway genes and proteins may be used
 CC to identify genes which are part of the mammalian pathway and
 CC to identify genes, proteins and therapeutic compounds which
 CC modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 CC C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 CC -33, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 CC acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 CC polypeptide, and (5) an antibody which binds to a SynMuv family
 CC protein. The SynMuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions
 CC involving altered levels of cell proliferation, e.g. SynMuv-
 CC associated carcinomas.
 XX SQ Sequence 438 AA;
 Alignment Scores:
 Pred. No.: 1,14e-33 Length: 438
 Score: 441.00 Matches: 93
 Percent Similarity: 55.31% Conservative: 32
 Best Local Similarity: 41.15% Mismatches: 61
 Query Match: 11.53% Indels: 40
 DB: 20 Gaps: 9
 US-09-743-237-3 (1-2134) x AAW83392 (1-438)
 Qy 698 TACTGTGACTGCTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTGTTGC 757
 Db 191 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 210
 Qy 758 AACAACTTGTCATCATGATATTGAACGGTTTAAACCCATTAAAGCATGCTCTTTGGTAGAAAT 817
 Db 211 AsnAsnIleGluTyrcysSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 230
 Qy 818 CCAGAAGCTTCCAGCCCAAAATTTGGG-----AAGGCCCAATTTGGCAATGTCACAGCC 871
 Db 231 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArg 250
 Qy 872 CAGCACAAACAAAGGTCACACTGCAGGAGTGCAGCTGCCTGGAAGAATTTACTGCGAGTGC 931
 Db 251 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrcysGluCys 270
 Qy 932 TATGAGCCCAAAATATTGTTCTTCTTATTTGCAAAATGCATTTGGTTCGCAAAAT----- 985
 Db 271 TyrGluAlaLysValProCysThrAspArgCysLysCysGlyCysGlnAsnThrGlu 290
 Qy 986 -----TATGAGAAGC-----CCAGACGAAGACACTA 1015
 Db 291 ThrTyrArgMetThrArgTyrcysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 310

Db 42 CysTyrGlnSerMetAlaIleCysThrLysPheCysArgCysValGlyCysArgAsnThr 61
QY 989 GAA-----GAAGCCCAACGA 1006
Db 62 GluValArgGluLeuValAspProAsnSerValAlaLysAsnSerSerAlaValLysArg 81
QY 1006 ----- 1006
Db 82 GlnLysAlaAlaMetSerAlaLysAlaAlaAlaAlaAlaAlaLysAlaGlyLeuAsp 101
QY 1007 -----AAGACACTA 1015
Db 102 ValGlnGlyLysAlaLeuGlnValAlaAlaSerThrLeuAlaLeuProGlyLysAlaLeu 121
QY 1016 ATGAGCATGCCAACTACATGCAGACTGGAGT-----TTGGAAGGCGCCATTAC 1066
Db 122 MetThrProProLysTyrThrLeuValAlaGlyLysProPheMetAlaSerSerHisLeu 141
QY 1067 CTGCCA-----CCACGAAATTTTCAGGACTTCCAGATTCAGT----- 1105
Db 142 AsnProIleProIleSerArgProIleAlaThrAlaAlaThrProAlaAlaValLys 161
QY 1106 -----CACCATAGG 1114
Db 162 GlnProAlaGluProProMetProValAsnLeuIleIleProValArgHisAspArg 181
QY 1115 CGGCCTTCTCATGCTCTCTGGCAGGTGGTGAGGCCACATGCCTGCTGCTGCT 1174
Db 182 ArgAspArgAsnLeuPheValGlnProValAsnAlaAlaLeuLeuGluCysMetLeuLeu 201
QY 1175 CAGGAGAGAGCGCGAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCCTG 1234
Db 202 GlnAlaThrGluAlaGluGlnLeuGluLeuAsnGluLeuGlnValCysGlnLeuValLeu 221
QY 1235 GAGGAATTGGAGGTGCTTATCAGATTCTC 1267
Db 222 GluGluPheMetArgGlyTyrLysAsnIleLeu 232

RESULT 11

ABG30203

ID ABG30203 standard; Protein; 4561 AA.

XX ABG30203;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #30194.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS94390.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity

XX Claim 20; SEQ ID No 60562; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 4561 AA;

Alignment Scores:

Pred. No.:	7,41e-05	Length:	4561
Score:	149.00	Matches:	157
Percent Similarity:	28.98%	Conservative:	54
Best Local Similarity:	21.57%	Mismatches:	205
Query Match:	3.90%	Indels:	312
DB:	22	Gaps:	45

US-09-743-237-3 (1-2134) x ABG30203 (1-4561)

QY 8 GGTCAAGCGAAGCTCGCGGGGCGACAGCGCGGGAGCTCT-----CGG 58
Db 1647 GlySerGlyArgCysSerAlaGlyArgProGlyLeuProGlyProAlaProLysAla 1666
QY 59 GGAGTACCCCGGATCCAG----- 79
Db 1667 AlaLeuProArgGluGlyAlaLeuAlaGlyAlaValProAspSerAlaProAlaLeu 1686
QY 80 -----GCTCAGCGCGCTGGAGGAGCTCGCGCTCTGCGAGGCCCGCCGCCGCC 127
Db 1687 GlyIleHisProAlaGlnGluMetProProArgSerProAlaAlaProLeuTyrAla 1706
QY 128 CGCTTGCAACGTGCAC-----TCCTGTCTCT----- 153
Db 1707 GlnIleGlu--CysThrGlyPheCysAlaProGlyCysThrCysProGlyLeuPheL 1726
QY 154 -----CGCTGTACCGCGCGCGCGAGCGCGGGGT----- 186
Db 1726 euHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeuHisGlyGlnL 1745
QY 187 -TTTGGCCCTGGCGCGCTGGG-----TCTGCGAAGGAGCGCTCCACCGCGCGG 235
Db 1745 eutyAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysCysAsnValGlyI 1765
QY 236 TCCGCATGATCCAGTTGAAATCAAGGTAAGCAGGTGTACTTACT----- 280
Db 1765 leArgArg-----ArgPhe-ArgAlaGlyThrAlaProAlaAlaPhe 1779
QY 281 -----ACAAGTATATCCGGAAGAGCAACTTTGCAGATCTT 319
Db 1780 GlyGlyAlaGluCysGlnGlyProThrMetGluAlaGluPheCysSerLeuArgProCys 1799
QY 320 CTTGTCTCAGGAATCCTGTTGC---AAGTTCCTCATGTCCCGAGAACTAGAGGATCCCTCC 376
Db 1800 ProAlaSerTrpTrpGlyCysHisArgValProCysThrGlyGluLeuGlu----- 1816

QY 377 TGCTGTTCTCTTAAGAAGATTCCACCCCAATGGTGTATATGCCAATTTGAAGGGGGCACA 436
 Db 1817 -----GlnArgProLeuMetValSerGlnIle----- 1825
 QY 437 CAAATGCTATGTATAGACAATTCTAGAACAGAGAACTAAAGCACTCCATTGGTTCT 496
 Db 1826 -----LeuGluAla----- 1828
 QY 497 CAGTATCAAGATCAAAATAATTATCTACAGTCAGATGCTCCCTAAACCAATGACTGCT 553
 Db 1829 -----GlnAspGlnGlyValAlaProValSerProGlySerProThrAlaAlaPro 1846
 QY 554 -----TTAGTAGGAGATTTTGCAGCATCAACAAAATAAATCTC-----ATTACACACAA 607
 Db 1847 GlyLeuValArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGln 1866
 QY 608 CTTGAGGAGCCTTACCATTACCGTAGTCAACGGGTCTCTTCCCTCGGGATCA----- 661
 Db 1867 ProProValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyTyr 1886
 QY 661 ----- 661
 Db 1887 GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg 1906
 QY 662 -----ACTCTTCAGACACACAAAATAACTTTGCTGGTACTGTGAC-----TGCTTT 712
 Db 1907 ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGluGlyProArg 1926
 QY 713 GCAGTGGGACTTTTGC-----AACACTGCAAT----- 742
 Db 1927 AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGlu 1946
 QY 743 -----TGTAATAAATTTGTTGCAACAACTTGCATCAT 772
 Db 1947 GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis 1966
 QY 773 GATATTGAACGGTTAAAGCCATTAGCATGCTCTGGTAGA-----AATCCAGAGCTTTC 829
 Db 1967 -----CysValTrpArgCysGlnProGlyCysTyr 1976
 QY 830 CAGCAAAAATTGGAGAGGCCAATTGGCAAT-----GTCAAGCCC 871
 Db 1977 CysPro-----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnPro 1993
 QY 872 CAGCAACAAAGGGTGCACACTGC----- 895
 Db 1994 GlyHis-----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGly 2010
 QY 896 -----AGGAGGTGAGCTGCCTGCAAGAAATTACTGCAGTCTATGAGGCCCAA 943
 Db 2011 AlaArgLeuAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArg 2028
 QY 944 ATTATGTGTTCTCTATTGCAAAATGATTGGTTGCAAAAATTATGAAGAAAGCCAGAA 1003
 Db 2029 LeuAsnCysThrAspLeu-----ProCysProAspCys----- 2039
 QY 1004 CGAAGACACAAATGACATGCCAAACTACATGCAGACTGGAGTTTGAAGGCAGCCAT 1063
 Db 2040 -----GlyGlyGlnSerLeuHis 2046
 QY 1064 TACCTGCCCAACGAAATTTTCAGACCTTCCAGATTCAGATTCAGTACGATAGCGGCTTCC 1123
 Db 2047 -----ProCysGly-----GlnProCysProArgSerCysGlnAspLeuSerProGly 2062
 QY 1124 TCA-----TGATCTCTCTGGGAGGTG-----GTGGAGGCCACATGCGCTGCTGTCTCAG 1177
 Db 2063 SerValCysGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln 2082
 QY 1178 -----GGAGAAGAGCCGAGAAAGAACACTGTCTCAAGTGCCTGGCAGACAG 1225
 Db 2083 LeuSerGlnAspGlyLeuCysValProProAlaHisCys----- 2095
 QY 1226 ATGATCTGGAGGAATTTGGAAGGTCTTATCACAG-----ATTCCTCCAC 1270

Db 2096 -----ArgCysGlnTyrGlnProGlyAlaMetGlyIleProGlu 2108
 QY 1271 ACTGAGTTTAAATCTAAGGAGTGAATAATGGAG-----TAGAGTATAAAG----- 1315
 Db 2109 AsnGlnSerArgSerAlaGlySerArgPheSerSerTrpGluSerLeuGluProGlyGlu 2128
 QY 1316 -----TGTGAATGCATGTTGATTTTGTCTTGTAGCTAGAAAT 1351
 Db 2129 ValValThrGlyProCysAspAsnCysThrCysValAlaGlyIle----- 2143
 QY 1352 CTCTAGTTTAGAAGGATTTTAGGGGAACATGAGGCTGCTCTGCAGCAACACACAGGC 1411
 Db 2144 -----LeuGlnCysGlnGluValProAsp 2151
 QY 1412 TCCCT-----GCATCCCTG 1426
 Db 2152 CysProAspProGlyValTrpSerSerTrpGlyProTrpGluAspCysSerValSerCys 2171
 QY 1427 GGGCCAGGAGTCTTACTCAGAGCTCTCTGAAGATGTGGCAAC-----CCATGCCCTTTTC 1483
 Db 2172 GlyGlyGlyGluGlnLeuArgSerArg-----ArgCysAlaArgProCysPro----- 2188
 QY 1484 TGAGGAGGTGCATGGCTGAGCATTTGTTCTGGCCAGAGAGAGAGCTTGGTTTCC 1543
 Db 2189 ---GlyProAla-ArgGlnSerArgThrCysSerThrGln----- 2200
 QY 1544 ATAGTCTGGGAGAGTCTCTGCAGGGCGGCGGAGGAGCAGCAGGC----- 1589
 Db 2201 -----ValCysArgGluAlaGlyCysProAlaGlyArgLeuTyrArgG1 2215
 QY 1590 -----CCTCGGAGAGCTCACTCTGGTCTCTCTCTCAGAGAA-----TG 1633
 Db 2215 uCysGlnProGlyGluGlyCysProPheSerCysAlaHisValThrGlnGlnValGlyC 2235
 QY 1634 TTCTCTGGAGGCTGC 1649
 Db 2235 sPheSerGluGlyCys 2240
 RESULT 12
 ABG21064
 ID ABG21064 standard; Protein; 9222 AA.
 AC ABG21064;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #21055.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS85251.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess

QY 1226 ATGATCTGGAGGAATTTGGAAGTGTCTTATCAG-----ATTCTCCAC 1270
 Db 2820 -----ArgCysGlnTrpGlnProGlyAlaMetGlyIleProGlu 2832
 QY 1271 ACTGAGTTAAATCAGGATTGAAATCGAG-----TAGAGTATAAG----- 1315
 Db 2833 AsnGlnSerArgSerAlaGlySerArgPheSerSerTrpGluSerLeuGluProGlyGlu 2852
 QY 1316 -----TGTGAATGCATGTGATTGTTGCTTGTAGTACAAAT 1351
 Db 2853 ValValThrGlyProCysAspAsnCysThrCysValAlaGlyIle----- 2867
 QY 1352 CTCTAGTTAGAAAGGATTTAGGGGAACATGAGCTGCTGCAGCAACACAGGC 1411
 Db 2868 -----LeuGlnCysGlnGluValProAsp 2875
 QY 1412 TCCCT-----GCATCCCTG 1426
 Db 2876 CysProAspProGlyValTrpSerSerTrpGlyProTrpGluAspCysSerValSerCys 2895
 QY 1427 GGCCCAAGGAGTTACTCAGAGCTCTCTGAAGATGTGGCAAC---CCATGCCCTTTTC 1483
 Db 2896 GlyGlyGlyGluGlnLeuArgSerArg---ArgCysAlaArgProProCysPro----- 2912
 QY 1484 TGAGGAGTGCATGGCTGAGCATTTCTTCTGCGCCAGAGAGAGAGCTGGGTCCC 1543
 Db 2913 ---GlyProAla-ArgGlnSerArgThrCysSerThrGln----- 2924
 QY 1544 ATAGTCTGGGAGAGTGTCTGCGAGCGCGGAGGCGAGCAGGC----- 1589
 Db 2925 -----ValCysArgGluAlaGlyCysProAlaGlyArgLeuTrpArgGln 1939
 QY 1590 -----CCTGCGAGAGCTCACTCTGTCGACTCTCTCTCTCAGAGAA-----TC 1633
 Db 2939 uCysGlnProGlyGluGlyCysProPheSerCysAlaHisValThrGlnGlnValGlyCys 2959
 QY 1634 TTCTCTGGAGGCTGC 1649
 Db 2959 sPheSerGluGlyCys 2964
 RESULT 13
 ABB81598
 ID ABB81598 standard; Protein; 2743 AA.
 AC ABB81598;
 XX
 XX
 19-SEP-2002 (first entry)
 XX Human laminin alpha 5 2743 N-terminal amino acid sequence SEQ ID NO:36.
 DE
 XX Laminin alpha 5; laminin 10; vulnerary; cell growth; differentiation;
 KW tissue repair development; laminin; healing; vascular tissue;
 KW re-endothelialisation; vascular injury; cell attachment; cell stasis;
 KW proliferation; migration.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= signal
 FT Protein 36..2743
 FT /label= laminin_alpha_5
 XX
 PN WO200250111-A2.
 XX
 XX 27-JUN-2002.
 PD
 XX 21-DEC-2001; 2001WO-US51035.
 PF
 XX 21-DEC-2000; 2000US-257449P.
 PR 28-MAR-2001; 2001US-279282P.
 PR 13-NOV-2001; 2001US-0279282.
 XX

(BIOS-) BIOSTRATUM INC.

Tryggvason K, Doi M, Thyboll J;

WPI; 2002-557650/59.

N-PSDB; ABQ72930.

New human laminin-10 proteins, useful for accelerating the healing of vascular tissue, improving the biocompatibility of grafts, or for promoting re-endothelialization at the site of vascular injuries

Disclosure; Page 223-231; 231pp; English.

The present invention describes human laminin alpha 5. Also described is an isolated laminin 10. Laminin 10 has vulnerary activity. Laminins are useful in maintaining cell/tissue phenotype as well as promoting cell growth and differentiation in tissue repair development. Specifically, laminin 10 can be used for accelerating the healing of injuries of vascular tissue, improving the biocompatibility of grafts useful for treating such injuries, for promoting re-endothelialisation at the site of vascular injuries, and promote cell attachment and subsequent cell stasis, proliferation, differentiation, and/or migration. The present sequence represents the 2743 N-terminal amino acid sequence of human laminin alpha 5, which is used in the exemplification of the present invention.

SQ Sequence 2743 AA;

Alignment Scores:
 Pred. No.: 0.00373 Length: 2743
 Score: 130.50 Matches: 111
 Percent Similarity: 31.14% Conservative: 55
 Best Local Similarity: 20.83% Mismatches: 198
 Query Match: 3.41% Indels: 170
 DB: 23 Gaps: 28

US-09-743-237-3 (1-2134) x ABB81598 (1-2743)

QY 18 AAGCTCGCGGGGCGACAGCGACGGCGGAG-----CTCCTCGGGAGTAC 65

Db 83 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr 102

QY 66 CCCGGGATC-----CCAGAGCTCAGCGCGCTGGAG 95

Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaLeuAsp 122

QY 96 GACGTC-----GGCTCTCGACGCCCGCGAGCGCGCTGC 134

Db 123 GlyThrGluArgTrpTrpGlnSerProLeuSerArgGlyLeuGluTrpAsnGluVal 142

QY 135 AACGTG-----CACTTCTCTGCTCGCTGCTACCCGCGCAC 170

Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162

QY 171 CGCAGCGCGGGTGTGTTTCCCTGGCGCGCTGGTCTCGGAGGAGCGCTCCACCC 230

Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178

QY 231 GGGCGTCCGCATGATCCAGTTGAAATCAAGTAAAGCAGGTGGTACTACTACAAGTAATA 290

Db 179 GlyArgThrTrpGlnPro-----TrpGlnPhePhe 188

QY 291 ATCCGGAAGAACAACTTTGCAAGATCTTTGCTCAGGAATCTCTGTTGCAAGTTCCCAT 350

Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199

QY 351 GGTCCCGAGACTA-----GAGGATGCTCTGCTGCTTCTTCTTAAAGAA 394

Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaAlaIleCysThrThrGluTyr 219

QY 395 GATTCACCAACCAATGTTGATATGCCAATTGAAAGGGGCGACACAATGCTGTATAGAC 454

Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValSerLeuVal 235

QY	66	CCGGGATC-----	-----CCAGAGCTCAGCCGCTGGAG	95
Db	103	CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp	AsnAlaIleAsp	122
QY	96	GACGTC-----	-----GGCCTCCTCAGCGCCGCGAGCGCCGCGCTGC	134
Db	123	GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal	AsnGluVal	142
QY	135	AACGTG-----	-----CACTTCCTCTCCTCGCTGCTACCCCGCGCAC	170
Db	143	AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla	LysPheAla	162
QY	171	CGCAGCCCGGGGTGTTTCCCTGGCGCCTCGGTCTCGAAGGAGCCTCCACCC	230	
Db	163	AsnSerProArgPro-----	-----AspLeuTrpValLeuGluArgSerMetAspPhe	178
QY	231	GGCGTCGCGCATGATCCAGTTGAATCAAGGTAAAGCAGGTGTACTACTACAAGTAATA	290	
Db	179	GlyArgThrTyrGlnPro-----	-----TrpGlnPhePhe	188
QY	291	ATCCGGAAGACCAAGCATTTGCAGAACTTCTTGTCTCAGGAATCCGTTCGAAGTTCCCAT	350	
Db	189	AlaSerSerLysArgAspCys-----	-----LeuGluArgPhe	199
QY	351	GGTCCAGGAACATA-----	-----GAGGATGCCTCGTGTCTTCTTAAAGAA	394
Db	200	GlyProGlnThrLeuGluArgIleThrArgAspAlaAlaIleCysThrThrGluTyr	219	
QY	395	GATTCCAACCAATGTTGATATGCCAATTAAGAGGGGCACAAATGCTATGTATAGAC	454	
Db	220	SerArg-----	-----IleValProLeuGluAsnGlyIleValValSerLeuVal	235
QY	455	AATTCTAGAACAGAGAACTTAAAGCACTCCATTTGGTTCTCAGTATCAAGATCAAAAT	514	
Db	236	AsnGlyArgProGlyAlaMet-----	-----	242
QY	515	AATTATCTACAGTCAGATGCTCCTAAACAAATGACTGCTTTAGTAGGG-----	AGA	565
Db	243	AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg	262	
QY	566	TTTTTCCAGCATCAACAAATA-----	-----AATCTCATTTACACAACTTTGAGGAGCCCTTA	622
Db	263	PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp-----	280	
QY	623	CCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGGATCAACTTTCCAGGAGCACCAAAA	682	
Db	281	ProThrValThrArgArgTyrTyr-----	-----SerIleLysAsp	293
QY	683	ATAACTTTGGCTGGGTACTGTGACTGCTTTGCCAGTGGGACTTTTGC-----	730	
Db	294	IleSerIleGlyArgCysValCysHisGlyHisAlaAspAlaCysAspAlaLysAsp	313	
QY	731	-----	-----AACCAACTGC-----AATGT	745
Db	314	ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyThrCys	333	
QY	746	AATAATTGTTGCAACAACTTGATCATGATATTGAACGGTTTAAAGCCATTAAAGCA---	802	
Db	334	AspArgCysCysProGlyPheAsnGln-----	-----GlnProTrpLysProAlaThrAlaAsn	351
QY	803	-----	-----TGCTGTTGTAGA-----AAT	817
Db	352	SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp	371	
QY	818	CCAGAA-----	-----GCTTTCCAGCAAAAATTTGGGAAGGGCCAAATTTGGGCAAT	862
Db	372	ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGlyGly	391	
QY	863	GTC-----	-----AAGCCCCAGACACAAAGGGTCACTGCAGGAGGTGAGGTGCGCTG	913
Db	392	ValCysIleAspCysGlnHisHisThrAlaGlyValAsnCysGluArg-----	-----CysLeu	409
QY	914	AAGAAATTACTCGAGTGCTATGAGGCCCAAAATATCTGTTCTTCTATTATTGCAAAATGCATT	973	

Db	410	ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys	429
Qy	974	GGTTGCAGAAAAT-----TATGAAGAAGCCAGAACGAAAGACACACTA	1015
Db	430	AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys	447
Qy	1016	ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGGAAAGCG	1057
Db	448	TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly	467
Qy	1058	AGCCATTACCTGCCACCAAGAAATTTTCAGGACTTCCAAGATTCAGTCACGATAGGCGG	1117
Db	468	PheProSerCysTyrProThr-----ProSerSerAsnAspThrArg	482
Qy	1118	CCTTCCCTCATGCATCTCCTGGGAGGTGGTGGAGGCCACATGCGCTGCTCTCAG	1177
Db	483	GlucInValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln	502
Qy	1178	GGAGAAGAGGCCGAGGAAGAACACTCTCCAAAGTCGCTGGCAGACGACATGCTCGAG	1237
Db	503	GlyAsnAlaCysArgLysAsp-----Pro	510
Qy	1238	GAATTTGGGAAGGTCCTATCA-----CAGATTCTCCACACTGAGTTTAAATCT	1285
Db	511	ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla	530
Qy	1286	AAGGGATTGAATGGAGTAGACTATAAAGTGTGAATGC	1324
Db	531	ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys	543
RESULT	15		
AAEL17310			
ID	AAEL17310	standard; Protein; 3696 AA.	
XX	AAEL17310;		
XX			
XX	18-APR-2002	(first entry)	
XX			
XX	Human laminin alpha protein, sbg417005LAMININ_ALPHA #2.		
XX	Human; therapy; wound healing disorder; vaccine; cancer; infection;		
XX	autoimmune disorder; haematopoietic disorder; inflammation; arthritis;		
XX	Parkinson's disease; Huntington's chorea; schizophrenia; antiarrhythmic		
XX	multiple sclerosis; Alzheimer's disease; analgesic; cardiant; asthma;		
XX	ischaemia; stroke; AIDS; bone disease; atherosclerosis; brain disorder		
XX	depression; cardiovascular disease; myocardial infarction; renal failure		
XX	respiratory disease; liver disorder; Fanconi's syndrome; spleen disorder		
XX	type II diabetes mellitus; skeletal muscle disorder; immunosuppressive		
XX	hypersplenism; renal disease; hypoglycaemia; gastrointestinal disease;		
XX	neotrophic; cirrhosis; Hodgkin's disease; neuroleptic; antiinflammatory		
XX	haemostatic; vulnery; anticonvulsant; antirheumatic; neuroprotective;		
XX	nephrotropic; hypotensive; vasotrophic; cytostatic; cerebroprotective;		
XX	allergy; laminin alpha protein.		
OS	Homo sapiens.		
XX			
XX	WO200198342-A1.		
XX			
XX	27-DEC-2001.		
XX			
XX	22-JUN-2001; 2001WO-US19929.		
XX			
XX	22-JUN-2000; 2000US-213156P.		
XX			
XX	22-JUN-2000; 2000US-213161P.		
XX			
XX	(SMIK) SMITHKLINE BEECHAM CORP.		
PA	(SMIK) SMITHKLINE BEECHAM PLC.		
PA	(GLAX) GLAXO GROUP LTD.		
XX			
PI	Agarwal P, Cogswell JP, Kabnic KS, Lai Y, Martensen SA;		
PI	Murdock PR, Smith RF, Strum JC, Xiang Z, Xie Q, Rizni SK;		

DR WPI: 2002-139783/18.
XX N-PSDB; AAD27805.
PT Novel secreted and membrane-associated polypeptides and polynucleotides
PT useful for preventing, ameliorating or correcting dysfunction or
PT disease including diabetes, cancer, hypertension and growth
PT abnormalities
XX
PS Claim 1; Page 115-122; 138pp; English.
XX
CC The invention relates to secreted and membrane-associated polypeptides
CC and polynucleotides. The sequences of the invention are useful in
CC diagnostic assays for detecting diseases associated with inappropriate
CC activity or levels of these polynucleotides, and in identifying their
CC agonists and antagonists that are potentially useful in therapy. The
CC sequences of the invention are useful as vaccines for inducing
CC immunological response. The sequences of the invention are useful for
CC treating cancers, infections, autoimmune disorders, hematopoietic
CC disorders, wound healing disorders, cholesterol ester storage disease,
CC inflammation, congenital muscular dystrophy, junctional epidermolysis
CC bullosa, Parkinson's disease, Huntington's chorea, multiple sclerosis,
CC viral and bacterial infections, Alzheimer's disease, asthma, arthritis,
CC allergies, schizophrenia, sbg442445PROA-associated disorders,
CC septicemia, psoriasis, inflammatory bowel disease, transplant rejection,
CC graft versus host disease, ischemia, stroke, acute respiratory disease
CC syndrome, restenosis, brain injury, AIDS, bone diseases, atherosclerosis,
CC brain disorders including parasupranuclear palsy, myotonic dystrophy,
CC depression, anxiety disorders and sleep disorders, cardiovascular
CC diseases including congestive heart failure and myocardial infarction,
CC respiratory diseases including chronic obstructive pulmonary disease,
CC acute bronchitis and adult respiratory distress syndrome, liver disorders
CC including hypercholesterolemia, hypertriglyceridaemia, cirrhosis, viral
CC and non-viral hepatitis, type II diabetes mellitus, renal disease
CC including acute and chronic renal failure, glomerulonephritis, Fanconi's
CC syndrome, cystinuria, skeletal muscle disorders including hypoglycaemia
CC and tendinitis, gastrointestinal diseases including intestinal
CC obstruction and tropical sprue, spleen disorders including hypersplenism,
CC Hodgkin's disease and malignant lymphoma, testicular cancer, male
CC reproductive diseases including low testosterone and male infertility.
CC The present sequence is human laminin alpha protein.
XX
SQ Sequence 3696 AA;

Alignment Scores:
Pred. No.: 0.00424 Length: 3696
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
DB: 23 Gaps: 28

US-09-743-237-3 (1-2134) x AAEL17310 (1-3696)

QY	18	AGCTCGCGGGGCGACAGCGCGGGAG-----CTCCTCGGGAGTAC	65
DB	83	LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr	102
QY	66	CCCGGATC-----CCAGCTCAGCGCGCTGGAG	95
DB	103	CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp	122
QY	96	GACGTC-----GGCTCCTCGAGCGCGCGCGCGCTGC	134
DB	123	GlyThrGluArgTrpGlnSerProProLeuSerArgGlyLeuGlnTyrAsnGluVal	142
QY	135	ACGGT-----CACTTCCTCGCTCGCTACCGCGCAC	170
DB	143	AnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla	162
QY	171	CGCAGCGCGGGTGTGTTGCCCTGGCGCGCTGGTCTCGAAGAGAGCTCCACCC	230
DB	163	AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe	178

QY	231	GGCGCTCGCATGATCCCAAGTAAATCAAGGTAAAGCGTGTCTACTACTACAAGTAATA	290
DB	179	GlyArgThrGlnPro-----TrpGlnPhePhe	188
QY	291	ATCCGGAAGCAAGCAATTTTGCAGATCTTCTTCTCCTCAGGAATCCTGTTCGAAGTCCCAT	350
DB	189	AlaSerSerLysArgAspCys-----LeuGluArgPhe	199
QY	351	GGTCCCGAGCACTA-----GAGGATGCTCTCTGTCTTCTTCTTAAGAA	394
DB	200	GlyProGlnThrLeuGluArgIleThrArgAspAlaAlaIleCysThrThrGluTyr	219
QY	395	GATTCACCAACCAATGATGATGCCAATTAAGGGGGGACACAAATGCTATATAGAC	454
DB	220	SerArg-----lleValProLeuGluAsnGlyGlyIleValValSerLeuVal	235
QY	455	AATTCTAGAACAGAACTAAAGCACTCCATTTGGTCTCCTCAGTATCAAGATCAAAAT	514
DB	236	AsnGlyArgProGlyAlaMet-----	242
QY	515	AATTATCTACAGTCAGATGCCCTAAACCAATCACTGCCTTAGTAGGG-----AGA	565
DB	243	AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg	262
QY	566	TTTTTCCAGCATCAACAAATTA---AATCTATTACAACTTGAGGGAGCCTTA	622
DB	263	PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp-----	280
QY	623	CCATCGGTAGTCAACGGGTCTGCTTCCCTCGGGATCAACTCTCCAGGACCAACAAA	682
DB	281	ProThrValThrArgArgTyrTyr-----SerIleLysAsp	293
QY	683	ATAACTTTGGCTGGGTACTGTGACTCTTCCAGTGGGACTTTTGC-----	730
DB	294	IleSerIleGlyGlyArgCysValCysHisGlyHisAlaAspAlaLysAsp	313
QY	731	-----AACAACTGC-----AATGT	745
DB	314	ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys	333
QY	746	AATAATTGTGCAACAACTTGCATCATATATGAAAGGTTTAAAGCCATTAAAGCA---	802
DB	334	AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn	351
QY	803	-----TGCTCTGGTAGA-----AAT	817
DB	352	SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp	371
QY	818	CCAGAA-----GCTTTCCAGCCCAAAATTTGGGAAGGCCCAATTTGGGCAAT	862
DB	372	ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGlyGly	391
QY	863	GTC-----AAGCCCCAGCACACAAGGTGCACTGCAGGAGGTGAGGCTGCCTG	913
DB	392	ValCysIleAspCysGlnHisThrThrGlyValAsnCysGluArg-----CysLeu	409
QY	914	AAGAAATTACTGCGAGTCTATGAGGCCCAATATGCTGTCTTCTTATTCGAATGCATT	973
DB	410	ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgCys	429
QY	974	GGTTGCAAAAT-----TATGAAGAAGCCCAAGCAAGACACTA	1015
DB	430	AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys	447
QY	1016	ATGAGCATGCCAACTAC-----ATGCAGACTGGAGTTTGAAGGC	1057
DB	448	TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly	467
QY	1058	AGCCATTACTGCCACCAAGAAATTTTCAGGACTTCCAGATTCACGATAGCGG	1117
DB	468	PheProSerCysTyrProThr-----ProSerSerAsnAspThrArg	482
QY	1118	CCTTCTCATCATCTCCTGGGAGGTGGTGGAGGCCACATGCGCTGCCTGTCTGCTCAG	1177

```
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysaspCysSerAlaAlaGlyThrGln 502
QY 1178 GGAGAAGAGCCCGAGAAGACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysAsp-----Pro 510
QY 1238 GAATTTGGAAGTGCTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
QY 1286 AAGGATTGAAAATGGAGTAGACTATATAAGTGTGAATGC 1324
Db 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543
```

Search completed: April 21, 2003, 11:55:41
Job time : 139.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:51:54 ; Search time 24.5 Seconds
(without alignments)
5125.593 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcgaagcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPro_spool/US09743237/runat_21042003_112211_1508/app_query.fasta_1.2311
-DB=Issued_Patents_AA -QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -CGN 1.1.33 -runat_21042003_112211_1508 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	3.2	739	4	US-08-444-818-148
2	124	3.2	2995	4	US-08-444-818-138
3	117	3.1	1021	1	US-07-910-760-12
4	117	3.1	1021	1	US-08-440-519-12
5	117	3.1	1021	4	US-08-440-549-12
6	117	3.1	2261	4	US-08-444-818-66
7	117	3.1	2436	4	US-08-444-818-75
8	117	3.1	2772	4	US-08-444-818-89
9	117	3.1	2894	2	US-08-466-975A-23
10	117	3.1	2894	2	US-08-391-671A-23
11	117	3.1	2894	3	US-08-467-902A-23
12	117	3.1	2894	4	US-09-275-265-23

13	117	3.1	3011	1	US-08-440-103-36	Sequence 36, Appl
14	117	3.1	3011	1	US-08-440-542-36	Sequence 36, Appl
15	117	3.1	3011	1	US-07-910-760-10	Sequence 10, Appl
16	117	3.1	3011	1	US-08-440-519-10	Sequence 10, Appl
17	117	3.1	3011	1	US-08-231-368-36	Sequence 36, Appl
18	117	3.1	3011	1	US-08-440-210-36	Sequence 36, Appl
19	117	3.1	3011	4	US-09-388-874-2	Sequence 2, Appl
20	117	3.1	3011	4	US-09-046-604-36	Sequence 36, Appl
21	117	3.1	3011	4	US-08-440-549-10	Sequence 10, Appl
22	117	3.1	3011	4	US-08-850-328-1	Sequence 1, Appl
23	116	3.0	1692	4	US-09-263-933-4	Sequence 4, Appl
24	116	3.0	1692	4	US-09-263-933-11	Sequence 11, Appl
25	116	3.0	1692	4	US-09-263-933-18	Sequence 18, Appl
26	116	3.0	2013	1	US-08-324-977-12	Sequence 12, Appl
27	116	3.0	2013	2	US-08-384-616-12	Sequence 12, Appl
28	116	3.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
29	116	3.0	2013	4	US-09-315-850-12	Sequence 12, Appl
30	116	3.0	2201	4	US-08-952-981A-2	Sequence 2, Appl
31	116	3.0	2307	4	US-09-263-933-2	Sequence 2, Appl
32	116	3.0	2307	4	US-09-263-933-9	Sequence 9, Appl
33	116	3.0	2307	4	US-09-263-933-16	Sequence 16, Appl
34	116	3.0	2620	1	US-08-324-977-32	Sequence 32, Appl
35	116	3.0	2620	2	US-08-384-616-32	Sequence 32, Appl
36	116	3.0	2620	4	US-08-904-686A-32	Sequence 32, Appl
37	116	3.0	2620	4	US-09-315-850-32	Sequence 32, Appl
38	116	3.0	2621	1	US-08-324-977-36	Sequence 36, Appl
39	116	3.0	2621	2	US-08-384-616-36	Sequence 36, Appl
40	116	3.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
41	116	3.0	2621	4	US-09-315-850-36	Sequence 36, Appl
42	116	3.0	3010	1	US-08-324-977-2	Sequence 2, Appl
43	116	3.0	3010	1	US-08-324-977-14	Sequence 14, Appl
44	116	3.0	3010	2	US-08-384-616-2	Sequence 2, Appl
45	116	3.0	3010	2	US-08-384-616-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-444-818-148
; Sequence 148, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-444-818-148

Alignment Scores:
 Pred. No.: 0.00155 Length: 739
 Score: 124.00 Matches: 91
 Percent Similarity: 35.32% Conservative: 45
 Best Local Similarity: 23.64% Mismatches: 140
 Query Match: 3.24% Indels: 109
 DB: 4 Gaps: 22

US-09-743-237-3 (1-2134) x US-08-444-818-148 (1-739)

```

QY 641 TCTGTTTCCCTCCGATCAACTCTCCAGGACCAACCAAAATAACTTTGGCTGGG 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 SerValIleProThrSerGlyAspValValValValValAlaThrAspAlaLeuMetThrGly 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 TACTGTCACTGCTTGGCACTGGGACTTTTGGCAACAACTGCAATTTGTAATAATTTGTC 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 AACAACTTG-----CATCATGATATTGACGGTTTAAAGCCATT 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 268 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrLeuPro 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 797 AAGGCATGCTTGTAGAAATCCAGAGCTTCCAGCCAAAATTTGGGAAGGCCCAATTG 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 857 GGCATGTCAAGCCCGCAGCACACAAA-----GGGTGCAACTGCAGAGGTCAAGCTGC 910
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GlyArgGlyLysProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 911 CTGAAGAAT-----TACTCGAGTGTATGAGGCCCAATTTATGTTCTTATTTCG 964
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 PheAspSerValLeuCysGluCysTyrAspAla----- 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 965 AAATGCATTGGTTCAGAAATATGAAGAAAGCCAGAGCAAGAAAGACACTAATGAGCATG 1024
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 -----GlyCysAlaTyrGluLeuThrPro--AlaGluThrThrValArgLeu 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1025 CCAAACTACATGACACTGAGGCTTG-----GAA 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1055 GGC-----AGCCATTACCTGCCCAACCAAGCAATTTTCA 1087
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1088 GGA-----CTTCCAAGATTTCATC-----ACGATAGGCGGCTTCTCTCATGC 1129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 GlyGluAsnLeuProThrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 ATCTCTGGGAGGTG-----TGGAGGCCACATGCGCTGCTGCTGCTCAGGCA 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 aProProSerTrpAspGlnMetTrpLys-----CysLeuIleArgLe 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1181 GAAGAGCGCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGACAGATGATCCTGGAGGAA 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 uLysProThrLeuHisGlyProThrProLeuTyrArgLeu----- 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1241 TTGGAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1278 -----TTAAATCTAAGGATTGAAATCGAGTAGAGTATAAGTGT 1318
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 tThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyGlyVa 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 1319 GAATGCA-----TGTTGATTTTCTCTAGTAATCTCTAGTTTAGAAGATGTT 1372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 lLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1373 TAGGGGAACATGAGCTGGCTCTCAGCAACACAGGCTCCCTCGCATCCCTGGGCCCA 1432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgG 508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1433 GGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTCTGA 1486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 uValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleG 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1487 GGAGGTGCATGGCTGAGCATTTGTTCTGTGGCCCAAGAGAGAGAGCTGGGTTCACATA 1546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 uGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu----- 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1547 GTCTGGGAGAGTGTCTGCGAGGGCGGAGGGGCGAGAGGCGCTCGGAGAGCTCACT 1606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnThrAs 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1607 CTGCTCGACTCTT 1619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 nTrpGlnLysLeu 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2

US-08-444-818-138
 ; Sequence 138, Application US/08444818
 ; Patent No. 6150087
 ; GENERAL INFORMATION:
 ; APPLICANT: Chien, David Y.
 ; APPLICANT: Rutter, William J.
 ; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
 ; NUMBER OF SEQUENCES: 777
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/444,818
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/403,590
 ; FILING DATE: 14-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Harbin, Alisa A.
 ; REGISTRATION NUMBER: 33,895
 ; REFERENCE/DOCKET NUMBER: 0110.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (508)359-3876
 ; TELEFAX: (508)359-3885
 ; INFORMATION FOR SEQ ID NO: 138:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2995 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-444-818-138

Alignment Scores:
 Pred. No.: 0.00286 Length: 2995
 Score: 124.00 Matches: 91
 Percent Similarity: 35.32% Conservative: 45
 Best Local Similarity: 23.64% Mismatches: 140
 Query Match: 3.24% Indels: 109


```
DB: 4 Gaps: 22
US-09-743-237-3 (1-2134) x US-08-444-818-138 (1-2995)
QY 641 TCTGTTTCCCC---TCGGATCAACTCTTCAGGACACCAAAATAACTTTGGCTGG 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGTTGCCAGTGGGACCTTTCCACAACTGCAATTAATAATTGTTGC 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
DB 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
QY 797 AAGCATGTCTGGTAGAATCCAGAAGCTTTCAGCCCAAAATTTGGGAAGGCCAAATTG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 1491
QY 857 GCAATGTCAAGCCCGACACACAAA-----GGTGCACCTGCAGGAGGTTCAGCTGC 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1492 GlyArgGlyProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 1511
QY 911 CTGAAGAAT-----TACTGCGAGTCTATGAGGCCAAATTTATGTGTTCTTCTATTGC 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1512 PheAspSerValLeuCysGluCysTyrAspAla----- 1523
QY 965 AATGCATTGTTGAAAATTTATGAAGAACCCAGACGAAAGACACTAATGACATG 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1524 -----GlyCysAlaTyrPyrGluLeuThrPro-----AlaGluThrThrValArgLeu 1539
QY 1025 CCAACTACACTCACACTGGAGTTG-----GAA 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1540 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 1559
QY 1055 GGC-----AGCCATTACTGCCACCAACGAAATTTTCA 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1560 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 1579
QY 1088 GGA-----CTTCCAAGATTACGTC-----ACGATAGGCGGCTTCCCTCATGC 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1580 GlyIleAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 1599
QY 1130 ATCTCTGGGAGTGG-----TGGAGGCCACATCGCTGCTGCTGCTCAGGGA 1180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1599 aProProSerTyrPaspGlnMetTrpLys-----CysLeuIleArgLeu 1614
QY 1181 GAAGAGCGCGAGAAACACTGCTCCAAAGTCCTGCGCAGCAGCAGATGATCCTGGAGGA 1240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1614 uLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1241 TTTGAAGTGTATTATCAGATTCTCCACTGAGT----- 1277
DB 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 1646
QY 1278 -----TTAAATCTAAGGATTGAAATGAGTACAGTATATAAGTCT 1318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1646 tThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyGlyVal 1666
QY 1319 GAATGCA-----TGTGATTGTTGTAGTCTAGAAATCTCTAGTTTAGAAGGATGTT 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1666 lLeuAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
QY 1373 TAGGGAACATGAGCTGCTGTCAGCAACCAACAGGCTCCCTGCAATCCCTGGGCCA 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgG1 1699
QY 1433 GGGAGTTTACTCAGAGCTCTCT-----GAAGATGGCGCAACCCATGCCCTTTCTGCA 1486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1699 uValLeuTyrArgGluPheAspGluMetGluCysSerGlnHisLeuProTyrIleG1 1719
QY 1487 GGAGGTGCATGGCTTGAGATTGTTTGTCTGGCCCGAGAGGAGAGCTTGGTTCCTCATA 1546
```

```

    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1719 uGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu----- 1735
QY 1547 GTCTGGGAGAGTGTCTCAGGCGCGGAGGCGAGAGCCCTCGCGAGAGCTCACT 1606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1736 ----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnThrAs 1753
QY 1607 CTGGTCGACTCTT 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1753 nTrpGlnLysLeu 1757

RESULT 3
; Sequence 12, Application US/07910760
; Patent No. 5683864
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/910,760
; FILING DATE: 07-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-910-760-12

Alignment Scores:
Pred. No.: 0.00863 Length: 1021
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x US-07-910-760-12 (1-1021)
QY 641 TCTGTTTCCCC---TCGGATCAACTCTTCAGGACACCAAAATAACTTTGGCTGG 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGTTGCCAGTGGGACCTTTGCAACAACACTGCAATTAATAATTGTTGC 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGGATGCTTGGTAGAATCCAGAGCTTCCAGCCAAAATTTGGAAGGCGCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGCCCGCAGCACACAAAGGTCGAAGGTCGAGGAGGTCAGGC 907
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAATTTATGTTCTTCTATT 961
Db 479 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 491
QY 962 TGCAAATGCATTGGTGTCAAAATTTATGAAGAACCCAGAACGAAACACACATATGAGC 1021
Db 492 -----GlyCysAlaTyrGluLeuThrPro-----AlaGluThrThrValArg 506
QY 1022 ATGCCAAACTACATCCAGCTGGAGTTG----- 1051
Db 507 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGGC-----AGCCATTACCTGCCACCAACGAAATTT 1084
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGGA-----CTTCCAGATTCAGTC-----ACGATAGCGGCTTCCTCA 1126
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCATCTCTCGGAGGTG-----TGGAGCCACATCGGCTCGCTGCTGCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 581
QY 1178 GGAGAAGAGCGGAGAAAGACACTGCTCCAAGTGCCTGGCAGCAGCATGATCCTGGAG 1237
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 596
QY 1238 GAATTGGAAGGTGCTTATCAGATTCACACTCAGT----- 1277
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 613
QY 1278 -----TTAAATCTAAGGATGAAATGGAGTAGAGTATTAAG 1315
Db 613 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyL1 633
QY 1316 TGCGAATGCA-----TGTTGATTGCTTAGTCTAGAAATCTCTAGTTTAAAGGAT 1369
Db 633 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGACATGAGCTGGCTCTCGACGACCAACAGCGTCCCTCGCATCCCTGGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 666
QY 1430 CCAGGGAGTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
Db 666 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 686
QY 1484 TGAGAGGTGCATGGCTGAGCATTTGTTGCTGGCCAGGAGGAGAGAGCTTGGGTTCC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCTGGAGAGTGTCTGCGAGGCGGGGAGGAGCAGAGCGCTCGGAGAGCTC 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla-----ProAlaValGlnTh 720
QY 1604 ACTCTGCTGACTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

```

RESULT 4

US-08-440-519-12

; Sequence 12, Application US/08440519

```

; Patent No. 5712087
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,519
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-519-12

```

```

Alignment Scores:
Pred. No.: 0.00863
Score: 117.00
Percent Similarity: 35.49%
Best Local Similarity: 23.58%
Query Match: 3.06%
DB: 111
Gaps: 23

```

US-09-743-237-3 (1-2134) x US-08-440-519-12 (1-1021)

```

QY 641 TCTGCTTCCCC---TCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 392 ServAlleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGCTTCCACTGGGACTTTTGCACCAACTGCAATTTGTAATAATTGTTGC 757
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACCAACTTG-----CATCATGATTTCAACGGTTTAAGCCATT 796
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGCATGCTTGGTGAATAATCCAGAGCTTCCAGCCAAAATTTGGAAGGCGCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGCCCGCAGCACACAAAGGTCGAAGGTCGAGGAGGTCAGGC 907
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAATTTATGTTCTTCTATT 961

```

```
Db 479 MetPheAspSerSerValLeuCysGluCysTyAspAla----- 491
QY 962 TGCAAAATGCTGGTTCAGAAATATGAAGAAACCCAGACGAAAGACACTAATGAGC 1021
Db 492 -----GlyCysAlaTrpTrpGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAATCACTACGACGACTGGAGTTTG----- 1051
Db 507 LeuArgAlaTyMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGC-----CTTCCAGATTTCAGTC-----ACGATAGGGGGCTTCCTCA 1126
Db 527 GluGlyValPheThrGlyLeuThrHisLeuAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGCA-----CTTCCAGATTTCAGTC-----ACGATAGGGGGCTTCCTCA 1126
Db 547 SerGlyGluAsnLeuProTyTrpLeu-ValAlaTyTrpGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCAATCTCTGGAGGTGG-----TGAGGGCCACATGGCGCTGCTGCTGCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 581
QY 1178 GGAGAGAGCGCGAAGAACACTGCTCCAAAGTCCCTGGCAGACAGATGATCCTGGAG 1237
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyArgLeu----- 596
QY 1238 GAATTTGGAAGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyTrl 613
QY 1278 -----TTAAATCTAAGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 613 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 633
QY 1316 TGTGAATGCA-----TGTGATTTTGTCTAGTCTAGTAATCTCTAGTTTGAAGAGAT 1369
Db 633 yValLeuAlaAlaLeuAlaAlaTyTrpCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGGAACATGAGCTGGCTGCGAGCAACACAGGCTCCCTGCGATCCCTGGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleProAspArg 666
QY 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCCCTTTC 1483
Db 666 gLluValLeuTyArgGluPheAspGluMetGluLucysSerGlnHisLeuProTyTrl 686
QY 1484 TGAGGAGTGCATGCCCTCAGCAATTTGTTGTGTCGCCAGAGGAGAGCTTGGGTTCCC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCCTGGGAGAGTGTCTGAGGGCGCGGAGGCGCAGACAGGCGCTGCGGAGAGCTC 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 720
QY 1604 ACTCTGGTCGAGCTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

RESULT 5
US-08-440-549-12
; Sequence 12, Application US/08440549
; Patent No. 6312889
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
```

```
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,549
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-549-12

Alignment Scores:
Pred. No.: 0.00863 Length: 1021
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23
```

US-09-743-237-3 (1-2134) x US-08-440-549-12 (1-1021)

```
QY 641 TCTGCTTTCCCC---TCGGGATCAACTCTTCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGCTTGGCCAGTGGGACTTTTGCACAACTGCAATTAATTAATTTGCTGC 757
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGGCATCTCTGGTAGAAATCCAGAGCTTTCCAGCCAAATAATTTGGGAAGGCCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGCCGCCAGCACAAAGGTCGCACTGCAGGAGGTGAGGC 907
Db 465 GlyIleTyArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATAGGCCCAATATGTGTTCTTCTATT 961
Db 479 MetPheAspSerSerValLeuCysGluCysTyAspAla----- 491
QY 962 TGCAAAATGCTGGTTCAGAAATATGAAGAAACCCAGACGAAAGACACTAATGAGC 1021
Db 492 -----GlyCysAlaTrpTrpGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAATCACTACGACGACTGGAGTTTG----- 1051
Db 507 LeuArgAlaTyMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
```



```
Db 988 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1003
QY 1238 GAATTTGGAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1004 -----GlyAlaValGlnAsnGluLeuThrLeuThrHisProValThrLysTyrI 1020
QY 1278 -----TTAAATCTAAGGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 1020 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG 1040
QY 1316 TGTGAATGCA-----TGTGTATTTGCTTCTAGTCTAGAAATCTCTAGTTTGAAGAGAT 1369
Db 1040 yValLeuAlaLeuAlaLeuAlaLeuSerThrGlyCysValValIleVal----- 1058
QY 1370 GTTTAGGGCAACATGAGTGGCTGCTGCAGCAACACAGGCTCCCTGCATCCCTGGGC 1429
Db 1059 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1073
QY 1430 CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
Db 1073 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 1093
QY 1484 TGAGGAGTGCATGGCTCAGCATTTGCTGTGCTGCGCCAGAGAGAGAGCTTGGGTCCC 1543
Db 1093 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1110
QY 1544 ATAGTCTGGGAGAGTGTGCGAGCGCGCGGAGGCGCAGCAGGCGCTTGGGAGAGCTC 1603
Db 1111 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1127
QY 1604 ACTCTGGTGCAGCTCTT 1619
Db 1127 rAsnTrpGlnLysLeu 1132
```

RESULT 7

```
US-08-444-818-75
; Sequence 75, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2436 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-75

Alignment Scores:
Pred. NO.: 0.0127 Length: 2436
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-444-818-75 (1-2436)
QY 641 TCTGCTTTCCCC--TCGGGATCACTCTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 974 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 993
QY 698 TACTGTGACTGCTTTGCCAGTGGGACTTTTGCACCAACTGCAATTTGTAATTTCTTGC 757
Db 994 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1008
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1009 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1028
QY 797 AGGGATGCTTCTGCTAGAAATCCAGNAGCTTCCAGCCCAAAATTTGGGAAGGCCAATTG 856
Db 1029 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1046
QY 857 GGC-----AATGTCAGGCCCCAGCAGCACAAAGGGTGCAACTGCAGGAGGTCAAGC 907
Db 1047 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1060
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGCTATGAGGCCCAATAATATGTCTTCTATT 961
Db 1061 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1073
QY 962 TCAATATGCTGTTGTCAGAAAATATGAAGAAAGCCAGAACGAAGACACATAATGAGC 1021
Db 1074 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1088
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTG----- 1051
Db 1089 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1108
QY 1052 GAAGGC-----AGCCATTACCTGCCAACACAGAAATTT 1084
Db 1109 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1128
QY 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGGCGGCTTCTCTCA 1126
Db 1129 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 1148
QY 1127 TCATCTCTCTGGGAGGTGG-----TGGAGGCCACATGCGCCCTGCTGTGCTCAG 1177
Db 1148 nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1163
QY 1178 GGAGAAGAGGCCGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATGATCTCTGGAG 1237
Db 1163 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1178
QY 1238 GAATTTGGAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1179 -----GlyAlaValGlnAsnGluLeuThrLeuThrHisProValThrLysTyrI 1195
QY 1278 -----TTAAATCTAAGGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 1195 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG 1215
QY 1316 TGTGAATGCA-----TGTGTATTTGCTTCTAGTCTAGAAATCTCTAGTTTGAAGAGAT 1369
Db 1111 -----TGTGTATTTGCTTCTAGTCTAGAAATCTCTAGTTTGAAGAGAT 1369
```


QY 1544 ATAGTCTGGAGAGTGTCTGCAGGCGGGAGGAGCAGAGCCCTCGGAGAGCTC 1603
Db 1622 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1638
QY 1604 ACTCTGGTCGACTCTT 1619
Db 1638 rAsnTrpGlnLysLeu 1643

RESULT 9

US-08-466-975A-23
; Sequence 23, Application US/08466975A
; Patent No. 5910404
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,975A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671
; FILING DATE:
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-466-975A-23

Alignment Scores:
Pred. No.: 0.0137 Length: 2894
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-466-975A-23 (1-2894)
QY 641 TCTGCTTCCCTCC---TCGGGATCAACTTCTCAGGACCAACCAAAATAAATCTTTGGTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGCTTTCCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATCTTTC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTCAACGGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrLeuPro 1478
QY 797 AAGGCATGCTTGGTAGAAATCCAGAGCTTCCAGCCCAAAATTTGGGAAGGCCCAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
QY 857 GGC-----AATGTCAAGCCCCAGCAGCACAAAGGGTGCACACTGCAGGAGTCAAGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
QY 908 TGCCTGAAGAAT-----TACTGCGAGTCTATGAGGCCCAAAATTTATGCTTCTTCTATT 961
Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
QY 962 TGCAAATGCTTGGTTGCAAAATTTATGAAGAAAGCCAGACGAACACATAATGAGC 1021
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1538
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
QY 1052 GAAGGC-----AGCCATTACCTGCCACCAAGAAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly 1578
QY 1085 TCAGGA-----CTTCCAAGATTCAGTC-----ACGATAGGCGGCTTCCCTCA 1126
Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaAlaArgAla 1598
QY 1127 TGCATCTCTCTGGGAGGTGG-----TGGAGGCCACATGCCCTGCTGCTGCTGCTCAG 1177
Db 1598 nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
QY 1178 GGAGAAGAGCGCGAGAAAGAACACTCTCCAGTGCCTGGCAGAGCAGATGATCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1238 GAATTTGGGAGGTGCTTATCAGAGATTCACACATG----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle 1645
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGATAAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly 1665
QY 1316 TGTGAATGCA-----TGTGTGATTTGCTTCTAGTCTAGAAATCTCTAGTTTAGAAAGAT 1369
Db 1665 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal----- 1683
QY 1370 GTTTAGGGGAACATGAGGCTGGCTCTGCAGCACACACAGGCTCCCTGCATCCCTGGGC 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
QY 1430 CCAGGAGGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATCTTCTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIle 1718
QY 1484 TGAGGAGGTGCATGGCCTGAGCATTTGTTCTGTGGCCCGCAGAGAGAGCTGGGTCCC 1543
Db 1484 ----- 1543

Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
Qy 1544 ATAGTCTGGAGAGTCTCTCAGGGCGGCGAGGCGAGAGCGGCTCGGAGAGCTC 1603
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
Qy 1604 ACTCTGGTGCAGCTTT 1619
Db 1752 rAsnTrpGlnLysLeu 1757

RESULT 10

US-08-391-671A-23
: Sequence 23, Application US/08391671A
: Patent No. 5922532
: GENERAL INFORMATION:
: APPLICANT: DELEYS, ROBERT J
: APPLICANT: POLLET, DIRK
: APPLICANT: MAERTENS, GEERT
: APPLICANT: VAN HEUVERSWUN, HUGO
: TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
: TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIXON & VANDERHIE P.C.
: STREET: 1100 NORTH GLEBE ROAD
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/391,671A
: FILING DATE: 21-FEB-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/920,286
: FILING DATE: 14-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/EP91/02409
: FILING DATE: 13-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 90124241.2
: FILING DATE: 14-DEC-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: SADOFF, B.J.
: REGISTRATION NUMBER: 36,663
: REFERENCE/DOCKET NUMBER: 1487-5
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 7038164000
: TELEFAX: 7038164100
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2894 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
US-08-391-671A-23

Alignment Scores:
Pred. No.: 0.0137 Length: 2894
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-391-671A-23 (1-2894)
Qy 641 TCTGCTTTCCCC---TCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
Qy 698 TACTGTGACTGCTTTCAGTGGGAGCTTTTGCACCAACTGCAATTTGTAATAATTTGTTC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
Qy 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
Qy 797 AAGCATGTCTTGTAGAAATCCAGAAGCTTTCAGGCAAAATTTGGGAGGCGCAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
Qy 857 GGC-----AATGTCAAGCCCGACGACACAAAGGGTGCAACTGCAGGAGGTGACGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
Qy 908 TGCCTGAAGAAT-----TACTGGGAGTGTATGAGGCCCAAAATATTGTGTCTTCTATT 961
Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
Qy 962 TGCAAAATGCATTGTGTCGCAAAATATTATGAAGAACCCAGAACAGACACATAATGAGC 1021
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro--AlaGluThrThrValArg 1538
Qy 1022 ATGCCAACTACATGCAGACTGGAGGTTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
Qy 1052 GAAGC-----AGCCATTACCTGCCACCACCAACGAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly 1578
Qy 1085 TCAGGA-----CTTCCAAGATTGAGTC-----ACGATAGCGGCGCTTCCTCA 1126
Db 1579 SerGlyGlnAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
Qy 1127 TGCATCTCTCTGGAGGTGG-----TGGAGGCCACATCGCGCTCGCTTGTCTGTCFAG 1177
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
Qy 1178 GGAGAGAGCCCGAGAGAACACTGCTCCAAGTCCCTGCGAGAGCAGATGATCCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
Qy 1238 GAATTTGGAAGGTGCTTATCACAGATTCTCCACACTCAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
Qy 1278 -----TTAAATCTAAGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 1665
Qy 1316 TGTCAATGCA-----TGTTGATTTTGTCTAGTCTAGAATCTCTAGTTTAGAAAGAT 1369
Db 1665 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
Qy 1370 GTTTAGGGGAACATGAGGCTGGCTGCAGCAACACAGCGTCCCTGCATCCCTGGG 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
Qy 1430 CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATCCCTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
Qy 1484 TGAGGAGTGCATGGCTGAGCATTGTTCTCGCCCGAGAGGAGAGCTGGTTCC 1543
Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735

RESULT 14

US-08-440-542-36
; Sequence 36, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/440,542

; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/231,368

; FILING DATE:

; APPLICATION NUMBER: US 07/759,575

; FILING DATE: 13-SEP-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: McClung, Barbara G.

; REGISTRATION NUMBER: 33,113

; REFERENCE/DOCKET NUMBER: 0205.001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (510) 601-2708

; TELEFAX: (510) 655-3542

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3011 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-440-542-36

Alignment Scores:
Pred. No.: 0.0139 Length: 3011
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-440-542-36 (1-3011)

```
Qy 641 TCTGTTTCCCC---TCGGGATCAACTCTCCAGGACCACCAAAATAAATTTGGCTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
Qy 698 TACTGTGACTGCTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATATTTGTC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
Qy 758 AACAACTTG-----CATCATGATATTTGAACGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
Qy 797 AAGGCATGCTTGGTGGAGAAATCCAGAACCTTCCAGCCCAAAATTTGGGAGGGCCAAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
```

```
Qy 857 GGC-----AATGTCAAGCCCCAGCAGCACAAACAAAGGTGCAACTGCAGGAGTCAAGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
Qy 908 TGCCTGAAGAAT-----TACTGGGAGTGTATGAGGCCCAAAATATGTGTTCTTATT 961
Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
Qy 962 TGCAATGATTTGGTTCACAAATATATGAAGAAAGCCAGACAGAAAGACACAACTAATGAGC 1021
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1538
Qy 1022 ATGCCAACTACATGCAGACTGGAGTTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
Qy 1052 GAAGGC-----AGCATTACCTGCCACCAACCAAAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1578
Qy 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGCGCGCTTCTCTCA 1126
Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
Qy 1127 TGCATCTCCTGGGAGTGG-----TGGAGGCCACATCGCGCTGCCTGCTGCTCAG 1177
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 1613
Qy 1178 GGAGAAGAGCGCCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
Qy 1238 GAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
Qy 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGTAGATATAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly1 1665
Qy 1316 TGTGAATGCA-----TGTTGATTTTGTCTAGTCTAGAAATCTCTAGTTTAGAAAGAT 1369
Db 1665 yValLeuAlaLeuAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
Qy 1370 GTTTAGGGGAACATGAGGCTGCTGCAGCACAAACACAGCTCCCTGCTGCTCCCTGGGC 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
Qy 1430 CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCCCTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
Qy 1484 TGAGGAGTGCATGCGCTGAGCATTTTGTCTGCGCCAGAGGAGAGAGCTTGGTGTCCC 1543
Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
Qy 1544 ATAGTCTGGGAGAGTGTCTGCAGGCGGGAGGAGGAGCAGCGCTCGCGGAGAGCTC 1603
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
Qy 1604 ACTCTGCTGACTCTT 1619
Db 1752 rAsnTrpGlnLysLeu 1757
```

RESULT 15

US-07-910-760-10
; Sequence 10, Application US/07910760
; Patent No. 5683864
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)

TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville
STATE: CA

COUNTRY: U.S.A.

ZIP: 94662-8097

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/910,760

FILING DATE: 07-JUL-1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Blackburn Esq., Robert P.

REGISTRATION NUMBER: 30,447

REFERENCE/DOCKET NUMBER: 0101.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2702

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-910-760-10

Alignment Scores:

Pred. No.:	0.0139	Length:	3011
Score:	117.00	Matches:	91
Percent Similarity:	35.49%	Conservative:	46
Best Local Similarity:	23.58%	Mismatches:	138
Query Match:	3.06%	Indels:	111
DB:	1	Gaps:	23

US-09-743-237-3 (1-2134) x US-07-910-760-10 (1-3011)

QY	641	TCTGCTTCCCC-TCGGGATCAACTCTCCAGGACCACCAAAATAACTTTGGCTGGG	697
DB	1424	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1478
QY	698	TACTGTGACTGTTTGGCCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTGTTGC	757
DB	1444	Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal	1458
QY	758	AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT	796
DB	1459	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1478
QY	797	AAGCATGCTTGTGGTAGAAATCCAGAAGCTTTCCAGCCAAAATAATGGGAAGGCCAAATG	856
DB	1479	GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlylysPro	1496
QY	857	GGC-----AATGTCAGCCCCAGCACAAAGGTCGCAACTGCAGGAGGTCAGGC	907
DB	1497	GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly	1510
QY	908	TGCCTGAAGAAAT-----TACTGCCAGTGCCTATGAGGCCCAATATATGTGTTCTTCTATT	961
DB	1511	MetPheAspSerSerValLeuCysGluCysTyrAspAla-----	1523
QY	962	TGCAAAATGCTTGGTGAATAATATGAAGAAAGCCAGACGACGAAAGACACTAATGAGC	1021
DB	1524	-----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg	1538
QY	1022	ATGCCAAACTACATGCAGCTGGAGTTG-----	1051

Search completed: April 21, 2003, 12:05:13
Job time : 105.5 secs

DB	1539	LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTyr	1558
QY	1052	GAAGC-----AGCCATTACCTGCCACCACGAATTT	1084
DB	1559	GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln	1578
QY	1085	TCAGGA-----CTTCCAAAGATTACATC-----ACGATAGGCGCTTCTCTCA	1126
DB	1579	SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGln	1598
QY	1127	TGCATCTCTCTGGGAGGTG-----TGGAGGCACATCGCCTGCTGCTGCTCAG	1177
DB	1598	nAlaProProSerTipAspGlnMetTriplys-----CysLeuIleAr	1613
QY	1178	GCAGAGAGGCGCAGAAAGACACATGCTCCAAAGTGCCTGGCAGACGATGATCTGGAG	1237
DB	1613	gLeuLysProThrLeuHisGlyProThrProLeuTyrArgLeu-----	1628
QY	1238	GAATTTGGAAGGTGCTTATCACAGATTCTCCACATGACT-----	1277
DB	1629	-----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle	1645
QY	1278	-----TTAAATCTAAGGATTGAAATGGAGTAGAGTATAAAG	1315
DB	1645	eMetThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyGln	1665
QY	1316	TGTGAATGCA-----TGTGATTTTGTCTTACTAGTAATCTCTAGTTTAAAGGAT	1369
DB	1665	yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValIleVal-----	1683
QY	1370	GTTTAGGGAACATGAGGCTGCTGCAGCAACCAACAGGCTCCCTGCATCCCTGGGC	1429
DB	1684	-----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr	1698
QY	1430	CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC	1483
DB	1698	gGluValLeuTyrArgGluPheAspGluMetGluCysSerGlnHisLeuProTyrIle	1718
QY	1484	TCAGGAGGTGCATGCGCTGAGCATTTGTTGCTGCGCCAGAGAGAGAGCTTGGTCCC	1543
DB	1718	eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu--	1735
QY	1544	ATAGTCTGGGAGAGTGTCTGCAGGCGCGGCGGAGGCGAGAGGCGCTTCCGAGAGCTC	1603
DB	1736	-----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh	1752
QY	1604	ACTCTGGTGCAGCTTT	1619
DB	1752	rAsnTrpGlnLysLeu	1757

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:55:45 ; Search time 53 seconds

(without alignments)
6089.058 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcggggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 577658

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_n2p.model -DEV=xlp
-Q=/cgn2_1/USProoool/US09743237/runat_21042003_112211_1541/app_query.fasta_1.2311
-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09743237 -CGCN 1.1.25 -runat_21042003_112211_1541
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	437	11.4	438	10	US-09-220-091-13
2	165	4.3	1057	9	US-10-189-971-6
3	158	4.1	1192	9	US-10-189-971-18
4	158	4.1	1207	9	US-10-189-971-20

5	158	4.1	1477	9	US-10-189-971-8	Sequence 8, Appl1
6	158	4.1	1535	9	US-10-189-971-14	Sequence 14, Appl
7	158	4.1	1593	9	US-10-189-971-4	Sequence 4, Appl1
8	157.5	4.1	1251	9	US-10-189-971-16	Sequence 16, Appl
9	157.5	4.1	1342	9	US-10-189-971-24	Sequence 24, Appl
10	157.5	4.1	1512	9	US-10-189-971-10	Sequence 10, Appl
11	157.5	4.1	1570	9	US-10-189-971-12	Sequence 12, Appl
12	157.5	4.1	1628	9	US-10-189-971-2	Sequence 2, Appl1
13	137	3.6	4123	9	US-10-213-509-5	Sequence 5, Appl1
14	130.5	3.4	2743	9	US-10-037-182-36	Sequence 36, Appl
15	130.5	3.4	3695	9	US-10-037-182-2	Sequence 2, Appl1
16	119	3.1	1036	9	US-09-887-527-60	Sequence 60, Appl
17	118.5	3.1	1801	10	US-09-938-275-8	Sequence 8, Appl1
18	118	3.1	732	9	US-09-977-418-20	Sequence 20, Appl
19	117	3.1	2894	9	US-10-044-995-23	Sequence 23, Appl
20	117	3.1	2894	10	US-09-941-611-23	Sequence 23, Appl
21	117	3.1	3011	10	US-09-916-359-2	Sequence 2, Appl1
22	117	3.1	3635	9	US-10-037-182-4	Sequence 4, Appl1
23	117	3.1	3635	10	US-09-845-583-2	Sequence 2, Appl1
24	116	3.0	1036	9	US-09-373-967-4	Sequence 4, Appl1
25	116	3.0	1036	9	US-10-028-072-142	Sequence 142, App
26	116	3.0	1036	9	US-10-121-049-142	Sequence 142, App
27	116	3.0	1036	9	US-10-123-904-142	Sequence 142, App
28	116	3.0	1036	9	US-10-140-470-142	Sequence 142, App
29	116	3.0	1036	9	US-10-175-746-142	Sequence 142, App
30	116	3.0	1036	9	US-10-176-918-142	Sequence 142, App
31	116	3.0	1036	9	US-10-176-921-142	Sequence 142, App
32	116	3.0	1036	9	US-10-137-863-142	Sequence 142, App
33	116	3.0	1036	9	US-10-140-474-142	Sequence 142, App
34	116	3.0	1036	9	US-10-142-431-142	Sequence 142, App
35	116	3.0	1036	9	US-10-143-114-142	Sequence 142, App
36	116	3.0	1036	9	US-10-140-002-142	Sequence 142, App
37	116	3.0	1036	9	US-10-142-419-142	Sequence 142, App
38	116	3.0	1036	9	US-10-123-262-142	Sequence 142, App
39	116	3.0	1036	9	US-10-142-423-142	Sequence 142, App
40	116	3.0	1036	9	US-10-121-050-142	Sequence 142, App
41	116	3.0	1036	9	US-10-141-755-142	Sequence 142, App
42	116	3.0	1036	9	US-10-143-032-142	Sequence 142, App
43	116	3.0	1036	9	US-10-123-108-142	Sequence 142, App
44	116	3.0	1036	9	US-10-123-236-142	Sequence 142, App
45	116	3.0	1036	9	US-10-123-261-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER FILING DATE: 1998-12-23
; EARLIER FILING DATE: 1997-05-28
; EARLIER FILING DATE: 1997-05-28
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13
Alignment Scores:
Pred. No.: 3.56e-29 Length: 438
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35

Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 10 Gaps: 8

US-09-743-237-3 (1-2134) x US-09-220-091-13 (1-438)

```
Qy 698 TACTGTGACTGCTTTGGCAGTGGGACTTTTGAACAACTGCAATTTGTAATAATTGTTGC 757
|||||
Db 191 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 210
|||||
Qy 758 AACCAACTGCATCATATTTGAACGCTTAAAGCCATTAAAGGCATCTCTGGTGAAGAAAT 817
|||||
Db 211 AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 230
|||||
Qy 818 CCAGAAAGCTTTCCAGCAAAATTTGGG-----AAGGCCAAATTTGGCAATGTCAAGCC 871
|||||
Db 231 ProAsnAlaPheLysProLysIleGlyLeuAlaArgGlyGlyIleThrAspIleGluArg 250
|||||
Qy 872 CAGCACAAAGGTCGCAACTCAGGAGGTGAGCTGCGCTGAAGAATTTACTGCGAGTGC 931
|||||
Db 251 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 270
|||||
Qy 932 TATCAGGCCCAAAATATGTTCTTCTATTGCAAAATGCAATGCTGTTGCCAAAAT----- 985
|||||
Db 271 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 290
|||||
Qy 986 -----TATGAAGAAAGC-----CCAGAACGAAGACACTA 1015
|||||
Db 291 ThrTyrArgMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 310
|||||
Qy 1016 ATGAGCATGCCAACTACATGCAGACT-----GGAGGTTTG--- 1051
|||||
Db 311 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 330
|||||
Qy 1052 -----GAAGCAGCCATTACTGCCACCACGAATTTTCAGACTTCCAGATTC 1102
|||||
Db 331 ThrAspGluHisGlyAspAspTyr---GluAspMetLeuLeuSerHisLysProLysVal 349
|||||
Qy 1103 AGTCACGATAGCGG-----CCVTCCTCATCATCTCCCTGGGAGGTGGTGGAGGCCACA 1156
|||||
Db 350 GluMetAspProArgArgPheProTyrTyrMetThrAspGluValValGluAlaAla 369
|||||
Qy 1157 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
|||||
Db 370 ThrMetCysMetValAlaGlnAlaGluGluAlaLeuAsnTyrGluLysValGlnThrGlu 389
|||||
Qy 1199 CACTGCTCCAGTCCCTGGCAGACAGATGATCCTGGAGGAATTTGGAAGGTGCTTATCA 1258
|||||
Db 390 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 409
|||||
Qy 1259 CAGATTCTC 1267
|||||
Db 410 GlnMetIle 412
```

RESULT 2

US-10-189-971-6

Sequence 6, Application US/10189971

Publication No. US20030028907A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Scoville, John

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleoti

TITLE OF INVENTION: Same

FILE REFERENCE: LEX-0360-USA

CURRENT APPLICATION NUMBER: US/10/189,971

PRIOR FILING DATE: 2002-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-08-29

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0.

SEQ ID NO 6
LENGTH: 1057
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-6

Alignment Scores:

Pred. No.: 1.33e-05 Length: 1057
Score: 165.00 Matches: 171
Percent Similarity: 27.67% Conservative: 62
Best Local Similarity: 20.31% Mismatches: 226
Query Match: 4.31% Indels: 383
DB: 9 Gaps: 48

US-09-743-237-3 (1-2134) x US-10-189-971-6 (1-1057)

```
Qy 16 CGAAGCTCGCGGGGGCGACAGCGCGGGGAGCTCCTCGGGGAGTACCCCGGGATCC 75
Db 24 ArgThrThrAlaAlaAlaValProLeuAlaGlyLys-----SerThrProAlaGlu 40
Qy 76 CAGAGCTTCAGCGCGCTGGAGAGC----- 99
Db 41 ArgThrSerProThrProLeuThrProAlaValCysValAlaValSerGlyAsnValGln 60
Qy 100 -----TCGCGCTCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138
Db 61 CysLeuAlaAlaArgArgCysValProLeuProCysProGluProValLeuLeuProGlyGlu 80
Qy 139 TGCACCTTCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 198
Db 81 Cys-----CysProGlnCysProAlaAlaProAlaGlyCysProArgProGly 98
Qy 199 GCG-----CCTGGGTCCTCGAAGGAGCGCTCC 225
Db 99 AlaAlaHisAlaArgHisGlnGluTyrPheSerProGlyAsp----- 113
Qy 226 CACCGCGGCTCGCATGATCCAGTTGAATCAAGTAAGCAGAGGTGGTACTACTACAAG 285
Db 114 ---ProCysArgArgCysLeuCysLeuAspGlySerValSer----- 126
Qy 286 TAATAATCCGGAAGAACCACTTTGCAAGATCTTCTGCTCAGGAATCCTGTTCAAGTT 345
Db 127 -----CysGlnArgLeuProCysProProAlaProCysAlaHis---ProArgGlnGly 143
Qy 346 CCCATGCTCCAGGAACATAGAGGATGCTCTG-----CTGTCTCTTAAGAA 393
Db 144 ProCysCysProSerCysAspGlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGlu 163
Qy 394 AGATTC---CAACCCAAATGTGATATGCCAAATTTGAAAGGGGGCACAAATATGCTATAT 450
Db 164 ArgPheProSerProThr-AlaAlaCysHisLeu-----CysLeuCysTr 178
Qy 451 AGACAATTCAGAACAGAGAACTTAAAGCACTC-----CATTTGCTTCTCT 496
Db 178 pGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCysProPheProAl 198
Qy 497 -----CAGTATCAAGATCAAAATATTTATCT 522
Db 198 aArgGlyAspCysCysProAspCysAspGlyCysGluTyrLeuGlyGlySerTyrLeuSe 218
Qy 523 ACAGTCAGATGCTCCTAAACCA-----ATGACTGCTTTAGTAGGAGATT 567
Db 218 rAsnGlnGluPheProAspProArgGluProCysAsnLeuCysThrCysLeuGlyGlyPh 238
Qy 568 TTTGCCAGCATCAACAAATTAATCTCATTACACAACTTGGAGGAGCGCTTACCATC 627
Db 238 eValThrCys-----GlyArgArgProCys 246
Qy 628 GGTAGTCAACGGTCTGCTTTTCCCC----- 652
Db 246 sGluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGl 266
Qy 653 -----TCGGGATCAACTCTTCCAGGACCA----- 676
```



```
QY 14 GCGAAGCTCCGGGGGCGACACGAGCGGGGGAGCTCTCGGGGAGTACCCCGGAT 73
Db 198 GlyCysAlaPheGlyGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 217
QY 74 CCCAG-----AGCTCAGCGCGCTGGAGGAGCTCGCGCTCTCGCAGG 114
Db 218 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 237
QY 115 CCCCGCAGCGCCCG-----CCTGCAACGTGCACCTCTCTCTCTCTCGCTGC 159
Db 238 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 255
QY 160 TACCCGCGCAGCGCCCGGGGTGTTGGCCCTCGGGCG-----201
Db 256 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 275
QY 202 -----CCTGGGTCTCGGAAGGAGCTCCACCGCGCGCTCCGATGATC 246
Db 276 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 289
QY 247 CCAGTTGAATCAAGGTAAAGCAGGTGTACTACTACAAGTAATAATCCGGAAGAACAAC 306
Db 290 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 301
QY 307 TTTCAGAAATCTTCTGCTCAGGAATCCTGTTCGAAGTCCCATGTTGCCAGGACTAGA 366
Db 302 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 320
QY 367 GGATGCTCTCTG-----CTGTCTCTTAAGAAAGATTC---CAACCCAAATGGT 411
Db 321 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL 340
QY 412 GATATGCCAATTAAGAGGGGCGCACAAATTCGTATATAGACAAATCTAGAACAAAGAGA 471
Db 340 aAlaCysHisLeu-----CysLeuCysTyrGluGlySerValSerCysGln 355
QY 472 ACTAAGACACTC-----CATTTGGTTCT-----496
Db 355 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 375
QY 497 -----CAGTATCAAGATCAAAATAATATTACTACATCGATGCTCCCTAAACC 543
Db 375 pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 395
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAAT 588
Db 395 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----411
QY 589 AAATCTCATTACACAACACTTCAGGAGCGCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 412 -----GlyArgArgProCysGluProProGlyCysSerH1 423
QY 649 CCCC-----652
Db 423 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 443
QY 653 -----TCGGGATCAACTCTCCAGACACCACCAAAATAACTTTGGCTGGGTACTG 702
Db 443 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr-----458
QY 703 TGACTGCTTTGCCAGTGGGACTTTTTCACAACTCAATTTGTAATTAATTTGTGCAACAA 762
Db 459 -----CysSerLeuCysThrCysGln-----GlyAr 467
QY 763 CTTGCATCATGATTAATTAAGCGGTTTAAAGCCATTAAGGCATGCTTGGTAGAAATCCAGA 822
Db 467 gGluHisGlnAspGlyGluPhe-----475
QY 823 AGCTTTCCAGCCAAAATTTGGGAGGGCCAAATTTGGCAATGTAAGCCCAAGCCCAACAA 882
Db 476 -----GluGlyProAlaGlySerCysGlu-----483
QY 883 AGSGTGCAACTGCAGGAGTCCGCTGCTGAGAAATTACTCGGAGTGTATGAGGCCCA 942
Db 484 -----TrpCysArgCysGlnAlaGlyG1 491
QY 943 AATTATGTGT-----952
Db 491 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 511
QY 953 -TCTTCTATTTCG---AAATGCATTGGTTGCAAAAATATAT---GAAGAAAGCCAGAACG 1005
Db 511 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro-----539
QY 1006 AAAGACACTAATAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGCGAGCCATTA 1065
Db 530 -----GluGlySerArgTr 534
QY 1066 CCTGCCCAACCAAAATTTTCAGGACTTCCAGATTCCAGATTCAGTCAGATAGGCGGCTTCCTC 1125
Db 534 pValProPro-----AspSerAlaCysSerSe 543
QY 1126 ATGCATCTCTCTGGAGGTGGTGGAGCCACATCGCG-----TGCTCG-----CT 1170
Db 543 rCysValCysHisGlyGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 562
QY 1171 TGCTCAGGAGAGAGCCGAGAAAGAACTGCTCCAAAGTGCCTGCCAGAGCAGATGAT 1230
Db 562 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys-----576
QY 1231 CCTGGAGGAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGAGTTAAATCTAAGGG 1290
Db 577 -----SerAspCysGluHisGluG1 583
QY 1291 ATTGAAAATGGAG-----TAGAGTATAAAGTGTCAATGCATGTT 1329
Db 583 yArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIl 603
QY 1330 GATTTGCTTAGTCTAGAAATCTAGTTTAGAAGAGTGTTTAGGGGAACATGAGGCT 1389
Db 603 e-----603
QY 1390 GGCTCTGTCAGACAACACAGGCTCCCTGCATCCCTGGGCGCCAGGAGTTTACTCAGAGC 1449
Db 604 ---CysGluProGlnProGluGlyProProSerLeu-----614
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCTTTTCTGAGGA-----1489
Db 615 ---ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 632
QY 1490 ---GGTGCATGCCCTGACATGT-----1510
Db 632 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 652
QY 1511 ---TTGCTGGCCCA-----GA 1524
Db 652 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGl 672
QY 1525 GGAGAGAGCTTGGTTTCCATA-----1546
Db 672 nasPLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuLeuSerCysProLeuSerGl 692
QY 1547 ---GTCCTGGGAGAGTGTCTCGAGGGCGCGG 1575
Db 692 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaAg 712
QY 1576 AGGGCAGACAGCCCTCGGAGAGCTCCTCTGCTCGACTCTCTCTCTCAGAGATGT- 1634
Db 712 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 730
QY 1635 ---TGCTCTGGAGCG-----TGCTCTGCATGAA 1659
Db 730 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla---748
QY 1660 AACCCTAATGGTTTCTTGTGTTTCTTCAATATTTAGAAATAGTTCTCGGATGGGC 1719
```

Db	749	-----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser	763
QY	1720	TGTTGTGATACCACCTTTAAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCGTAGC	1779
Db	764	:::-----CysGln	769
QY	1780	GTAGATATTTCCCA-----GAGACACGGGAACGTCTCAGTCTTTCCTTAAGCGCCCC	1830
Db	770	AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer	784
QY	1831	GGAGACAGCAGGCAATGGGGCCCTCCGAGGCCAGGGTTCGACACAGCATGCTCTTGATGTAGA	1890
Db	785	GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetala	800
QY	1891	GGACTTAAATATTCACGTTTCTCTGCGTTTCTACTTGAATTGTGGA	1938
Db	801	GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly	816

QY	247	CCAGTTGAATCAAGTAAGCAGGTGGTACTACTACAAAGTAATAATCCGGAAGAACAC	306
DB		
QY	305	CysLeuAspGlySerValSer	316
QY	307	TTTGCAGAACTCTTTCGTCAGGAATCCTGTTGCAAGTTCCCATGGTCCCGAGNACTAGA	366
DB		...	
QY	317	CysProProAlaProCysAlaHis	335
QY	367	GGATGCTCTCTG	411
DB			
QY	336	GlyCysLeuTyrlGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al	355
QY	412	GATATGCCAATTCGAAGGGGGCACAAATGCTATGTATAGACAATTTCTAGACAAGAGA	471
DB			
QY	355	aAlaCysHisLeu	370
QY	472	ACTAAAGCACTC	496
DB			
QY	370	uProLysAlaCysAlaProAlaLeuCyProPheProAlaArgGlyAspCysCysProAs	390
QY	497	-----CAGTATCAAGATCAAAATAAATATATCTACAGTCAGATGCTCCATAACC	543
DB			
QY	390	pCysAspGlyCysGluTyrlLeuGlySerTyrlLeuSerAsnGlnGluPheProAspPr	410
QY	544	A-----ATGACTGCTTTATAGTAGGAGATTTTGGCCAGCATCAACAATAAT	598
DB			
QY	410	oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys	426
QY	589	AAATCTCATTACACACACTTTGAGGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT	648
DB			
QY	427	-----GlyArgArgProCysGluProProGlyCysSerH	438
QY	649	CCCC	652
DB			
QY	438	sProLeuIleProSerGlyHisCysCyProThrCysGlnGlyCysArgTyrlHisGlyIva	458
QY	653	-----TCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTG	702
DB			
QY	458	lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr	473
QY	703	TGACTGCTTTCCAGTGGGACTTTTGCACACTGCAATTTGTAATTAATTTGTTGCAACAA	762
DB			
QY	474	-----CysSerLeuCysThrCysGln	482
QY	763	CTTGTCATCATGATTTGAACGGTTTAAAGCCATTAAGCATGCTCTGGTAGAATCCAGA	822
DB			
QY	482	gGluHisGlnAspGlyGluLuphe	490
QY	823	AGCTTTCCAGCCAAAATTTGGGAAGGGCCAAATTTGGGCAATGTCAGGCCCCAGCACACAA	882
DB			
QY	491	-----GluGlyProAlaGlySerCysGlu	498
QY	883	AGGTGCACTGCAGAGAGTCAGGTCGCTGAGCAATTTACTCGAGTGCTATGAGGCCCA	942
DB			
QY	499	-----TrpCysArgCysGlnAlaGlyG	506
QY	943	AATTATGTG	952
DB			
QY	506	nValSerCysValArgLeuGlnCyProProLeuProCysLysLeuGlnValThrGluAr	536
QY	953	-TCTTCTATTTCG- -AAATGCAATTGGTTGCAAAATTTAT- -GAAGAAAGCCCAAGAC	1005
DB			
QY	526	gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluHisPro	544
QY	1006	AAAGACACTAATGAGCATGCCAAACTACATGCAGACTCGGAGGTTTGGAAGCGACCCATTA	1065
DB			
QY	545	-----GluGlySerArgTr	549
QY	1066	CCTGCCACCAACGAATTTTCAGGACTTCCAGATTCCAGATTACATAGGCGGCTCTCCTC	1125
DB			
QY	549	pValProPro	558
QY	1126	ATGCATCTCTGGAGGTGGTGGAGCCACATGCGCC	1170
DB		-----TGCCTG-----	


```
QY 497 -----CAGTATCAAGATCAAAATAATATTATCTACAGTCAGATGTCCTCAACCC 543
Db 660 pCysAspGlyCysGluThrLeuGlyGluSerTyLeuSerAsnGlnGluPheProAspPr 680
QY 544 A-----ATGACTGCTTTAGTACGGAGATTTTCCAGCAGCATCAACAAAT 588
Db 680 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 696
QY 589 AAATCTCATACAAACACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 697 -----GlyArgArgProCysGluProProGlyCysSerHi 708
QY 649 CCCC----- 652
Db 708 sProLeuileProSerGlyHisCysCysProThrCysGlnGlyCysArgTyRHISGlyVa 728
QY 653 -----TCGGATCACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTG 702
Db 728 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 743
QY 703 TGAATGCTTTGCCAGTGGGACTTTTGCAACAACTGCAATTCATAATTTGTCACAA 762
Db 744 -----CysSerLeuCysThrCysGln-----GlyAr 752
QY 763 CTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATCTCTTGGTAGAAATCCAGA 822
Db 752 gGluHisGlnAspGlyGluGluPhe----- 760
QY 823 ACCTTTCCAGCCAAATAATTGGGAAGGCCAATTTGGCAATGTCAGAGCCCAACACAA 882
Db 761 -----GluGlyProAlaGlySerCysGlu----- 768
QY 883 AGGGTGCACTGCAGGAGTCAGGCTGCTGAAGAAATTAAGTCTCGAGTGTATGAGGCCCA 942
Db 769 -----irpCysArgCysGlnAlaGlyGly 776
QY 943 AATTATGTGT----- 952
Db 776 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 796
QY 953 -TCTTCTATTGCG---AAATGATTTGGTTGCAAAATAT---GAAGAAGCCCAAGAGC 1005
Db 796 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 814
QY 1006 AAAGACACTAATGAGCATGCCAACTACATGACAGCTGGAGGTTTGGAGGCGACCATTA 1065
Db 815 -----GluGlySerArgTr 819
QY 1066 CTGCGCACCAACAAATTTTCAGGACTTCCAAAGATTTCAGTCAGTAGCGGCTTCCTC 1125
Db 819 pValProPro-----AspSerAlaCysSerSe 828
QY 1126 ATGCATCTCTGGAGGTGGTGGAGGCCACATGCGGC-----TGCTG-----CT 1170
Db 828 rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 847
QY 1171 TGCTCAGGAGGAAGAGCGCGAGAAACACACTGCTCCAAAGTCCCTGGCAGACAGATGAT 1230
Db 847 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 861
QY 1231 CTGGAGGAATTTGGAAGGTGCTTATCAGAGATTCTCCACACTGAGTTAAATTAAGG 1290
Db 862 -----SerAspCysGluHisGluCl 868
QY 1291 ATTGAATGGAG-----TAGAGTATAAGTGTGAATGATGTT 1329
Db 868 yArgLysTyRGluproGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIl 888
QY 1330 GATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCT 1389
Db 888 e----- 888
QY 1390 GGCTCTGCAGCAACACCAGGCTCCCTGCTATCCCTGGGCCCGAGGAGTTTACTCAGAGC 1449
```

```
Db 889 -----CysGluProGlnProGluGlyProProSerLeu----- 899
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCCCTTTCTCAGGA----- 1489
Db 900 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 917
QY 1490 -----GGTGCATGGCTGAGCATTTG----- 1510
Db 917 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 937
QY 1511 -----TTGCTGTGGCCCA-----GA 1524
Db 937 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyThrCysGlnCysGl 957
QY 1525 GGAGAGAGCTTGGGTCCCAT----- 1546
Db 957 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGl 977
QY 1547 -----GTCTGGGAGAGTGTCTCGAGGCGCGG 1575
Db 977 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 997
QY 1576 AGGCGAGAGCAGGCGCTCGGAGAGCTCACTCTGGTCTGACTCTCTCTCAGAGAATGT- 1634
Db 997 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 1015
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
Db 1015 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla----- 1033
QY 1660 AACCTAATGTTTCTGTTCTTTTCAAAATTAATTTAGAAATTAAGTTCTCCGGATGGGC 1719
Db 1034 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 1048
QY 1720 TGTGTGTATACCACTAAATCTCTAGAGAACTACTGAACACCTTAAAGATTTCCTGATGC 1779
Db 1049 CysCysGluArg-----CysGln 1054
QY 1780 GTAGATATTTCCTCCA-----GAGACACCGAAGCTCTCAGTCTTCTTAAAGCCCCC 1830
Db 1055 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 1069
QY 1831 GGGAGCAGCAGCAATGGGCGCTCGCAGGCCAGGCTTGACACGAGCATGCTTTGAGTTAGA 1890
Db 1070 GlyGluArg-----TrpThrVal-----AspThrCysTrpSerCysSerCysMetAla 1085
QY 1891 GGACTTAAATTTATCCAGTTTCTCTGTGTTTCTTACTTGAATTTGTTGA 1938
Db 1086 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1101
```

RESULT 6

```
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302, 949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315, 634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
```

ORGANISM: homo sapiens
US-10-189-971-14

Alignment Scores:

Pred. No.: 6,09e-05 Length: 1535
Score: 158.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: 9 Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-14 (1-1535)

```
QY 14 GCGCAAGCTCGCGGGGCGGACGACGCGCGGAGTCTCTCGGGAGTACCCGGGAT 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GlyCysAlaPheGlyGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 74 CCCAG-----AGTCACGCGCGTGGAGAGCGTCCGCTCCTCGCAGG 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 CCCGCGAGCCCGG-----CCTGCAACGTGCGACTTCTCTCTCTCTCGTGC 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 TACCCGCGCACCGGCGCGGCGGTGTTTGCCTCGGCGG----- 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 -----CCTGGGCTCGGAAGAGCGCTCCACCGCGGCGTCCGCGATATC 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 619 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 CCAGTTGAATCAAGTAAGCAGGTGGTACTACTACAAAGTAATAATCCGGAAGAGCAAC 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 644
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 TTTGCAGAAATCTTCTCTCAGGAATCCTGTGTCAAGTTCCTCCAGTGGTCCAGGAACATAGA 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 CysProAlaProCysAlaHis--ProArgGlnGlyProCysCysProSerCysAsp 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 GGATGCTCTCTG-----CTGCTCTTAAAGAAATTC-----CAACCCCAATGGT 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 412 GATATGCCAATTGAAGGGGCGCACAAATGCTATGTATAGACAAATCTAGAACACAGAGA 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 AlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 698
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 472 ACTAAAGCACTC-----CATTTGGTTCCT----- 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 718
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 497 -----CAGTATCAAGATCAAAATAATATCTACGTACAGTACGATCCCTAAACC 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 544 A-----ATGACTGCTTTAGTACGAGATTTTTCGCCAGCATCAACAAAT 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 754
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 AAATCTATTACACAACAACTTGAAGGAGCGCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 755 -----GlyArgArgProCysGluProProGlyCysSerHis 766
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 CCCC----- 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 653 -----TCGGGATCAACTTTCAGGACACCAAAATAACTTTGCTGGGTACTG 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 lThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 801
```

```
QY 703 TGACTGCTTTGCCAGTGGGAGCTTTTGCAACAACCTGCAATTTGTAATAATTTGTCACAA 762
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 -----CysSerLeuCysThrCysGln-----GlyAr 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 763 CTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTCTGGTAGAATCCAGA 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 810 gGluHisGlnAspGlyGluGluPhe----- 818
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 823 AGCTTTCCAGCCCAAAATTTGGGAAGGCGCAATTTGGCAATGTCAAGCCGCCAGCACAA 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 819 -----GluGlyProAlaGlySerCysGlu----- 826
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 883 AGGTGCAACTGCAGGAGGTACAGCTCCCTGAAGAATTACTGCGAGTGTCTATGAGGCCCA 942
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 -----TipCysArgCysGlnAlaGlyGln 834
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 943 AATTATGTGT----- 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 834 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 953 -TCTTCTATTTCG---AATGCAATTGTTGTCACAAAATTAT---GAAGAAAGCCAGAACG 1005
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 872
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 AAAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGTTTGGAGGCGCACCATTA 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 -----GluGlySerArgTr 877
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1066 CTGGCCACCAACCAAAATTTTCAGGACTTCCAGATTCAGTCAGATAGGCGGCTTCCTC 1125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 pValProPro-----AspSerAlaCysSerSe 886
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 ATCATCTCTCTGGAGGTGGTGGAGGCCACATCGCCG-----TGCCTG-----CT 1170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 905
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1171 TGCTCAGGGAGAGAGCGCCGAGAAAGAACTGCTCCAGTGCCTGGCAGAGCAGATGAT 1230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 919
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 CCTGGAGGAATTTGGAGAGTGTATACAGATTCACAGATTCACACACTGAGTTTAAATCTAAGG 1290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 920 -----SerAspCysGluHisGluGln 926
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 ATTGAAAATGGAG-----TAGAGTATAAAGTGTCAATGCAATGCTT 1329
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 926 YArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysII 946
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1330 GATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCT 1389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 e----- 946
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1390 GGCTCTGCAGCAACACACAGGCTCCCTGCTGATCCCTGGGCGCCAGGAGTTTACTCAGAGC 1449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 ---CysGluProGlnProGluGlyProProSerLeu----- 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1450 TCTCTGAAGATGT---GGCAACCCATCGCCCTTTTCTGAGGA----- 1489
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 958 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 975
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1490 ---GGTGCATGCGCTGAGCATTTG----- 1510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 995
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1511 -----TTGCTGCCCA-----GA 1524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 995 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGln 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1525 GGAGAGAGCTTGGTTCGCCATA----- 1546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGln 1035
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1547 -----GTCCTGGGAGAGTGTCTGCAGGCGCGCGG 1575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1035 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 1055
||| |||||:|
QY 1576 AGGCGAGAGCCCTCGCGAGAGCTCACTCTGGTCGACTCTCTCTCAGAGATGT- 1634
||| |||||
Db 1055 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 1073
||| |||||
QY 1635 ---TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
||| |||||
Db 1073 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla---- 1091
||| |||||
QY 1660 AACCTAATGGTTCTGTTGTTGTTTTCAAATATTATTAGAAATAAGTCTCCGATGGC 1719
||| |||||
Db 1092 ---LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 1106
||| |||||
QY 1720 TGTGTGATACCACTTAAATCTCTAGAGAACTACTGAACACCTAAGATTTCTGTAGC 1779
||| |||||
Db 1107 CysCysGluArg-----CysGln 1112
||| |||||
QY 1780 GTAGATATTTCCCA-----GAGACACGCGAACTGTCACTTCTTCTAAGGCCCCC 1830
||| |||||
Db 1113 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 1127
||| |||||
QY 1831 GGGAGCGCAGGAATGGCGCCCTCGCAGCCAGCGCTTGCACACGATGTCTTGAGTTAGA 1890
||| |||||
Db 1128 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetAla 1143
||| |||||
QY 1891 GGACTTAAATATTCCAGTTTCTCTGTGTTCTTACTTGAATTGTGGA 1938
||| |||||
Db 1144 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1159
||| |||||

RESULT 7

US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleotides
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens

US-10-189-971-4

Alignment Scores:
Pred. No.: 6.16e-05 Length: 1593
Score: 159.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: 9 Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-4 (1-1593)

QY 14 GGGAGAGCTCGCGGGGCGGACACGACGCGGGAGCTCTCGGGAGGTACCCCGGAT 73
||| ||| |||||:|
Db 599 GlyCysAlaPheGlyGlyLysGluTyProSerGlyAlaAspPheProHisProSerAsp 618
||| ||| |||||
QY 74 CCCAG-----AGCTCAGCGCGTGGAGGAGCTCGCGCTCTCTGCAGG 114
||| ||| |||||
Db 619 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 638
||| ||| |||||

QY 115 CC CGCAGCCGCGCG-----CCTGCAACGTCCACTTCTCTCTCGCTGC 159
||| ||| |||||
Db 639 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 656
||| ||| |||||
QY 160 TACCCGCGCAGCCGCGCGGTGTTTCCCTCCCTGGGCG----- 201
||| ||| |||||
Db 657 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaAargHisGln 676
||| ||| |||||
QY 202 -----CCTGGGTCTCGAAGAGCCCTCCACCCGCGGCGTCCGATGATC 246
||| ||| |||||
Db 677 GluTyProPheSerProProGlyAsp-----ProCysArgArgCysLeu 690
||| ||| |||||
QY 247 CCAGTTCAATCAAGTAACGACGAGTGTACTACTACAAGTAATAATCCGGAAGAAGCAAC 306
||| ||| |||||
Db 691 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 702
||| ||| |||||
QY 307 TTTGCAAGATCTTCTTCTCTCAGGAATCCTGTTCAGACTTCCCATGTCCTCCAGGAACATAGA 366
||| ||| |||||
Db 703 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 721
||| ||| |||||
QY 367 GGATGCTCTCTG-----CTTCTCTTAAGAAAGATTC-----CAACCCCAATGGT 411
||| ||| |||||
Db 722 GlyCysLeuTyProGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 741
||| ||| |||||
QY 412 GATATGCCAATTTGAAAGGGGCGACACAATGCTATGTATAGACAATTTCTAGACAACAGA 471
||| ||| |||||
Db 741 alaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGly 756
||| ||| |||||
QY 472 ACTAAAGACACTC-----CATTTGGTTCCT----- 496
||| ||| |||||
Db 756 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 776
||| ||| |||||
QY 497 -----CAGTATCAAGATCAAAATATTATTCTACATCAGATGTCCTTAACC 543
||| ||| |||||
Db 776 pCysAspGlyCysGluTyProLeuGlyGluSerTyProLeuSerAsnGlnGluPheProAspPr 796
||| ||| |||||
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTGGCAGCATCAACAAAATT 588
||| ||| |||||
Db 796 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 812
||| ||| |||||
QY 589 AAATCTCATTTACACAACAACTTACGAGGAGCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
||| ||| |||||
Db 813 -----GlyArgArgProCysGluProProGlyCysSerH1 824
||| ||| |||||
QY 649 CCCC----- 652
||| ||| |||||
Db 824 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyHisGlyVa 844
||| ||| |||||
QY 653 -----TCGGGATCAACTCTTCCAGGACCCACCAAAATAACTTTGGCTGGGTACTG 702
||| ||| |||||
Db 844 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 859
||| ||| |||||
QY 703 TGACTGCTTTGCCAGTGGGAGCTTTTGCACAACAACTGCAATTTGTAATAATTGTTGCAACAA 762
||| ||| |||||
Db 860 -----CysSerLeuCysThrCysGln-----GlyAr 868
||| ||| |||||
QY 763 CTTGTCATCATGATTTCAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGA 822
||| ||| |||||
Db 868 gGluHisGlnAspGlyGluGluPhe----- 876
||| ||| |||||
QY 823 AGCTTTCCAGCCCAAAATTTGGGAAGGGCCCAATTTGGCAATGTCAAGCCCGCCAGCACAA 882
||| ||| |||||
Db 877 -----GluGlyProAlaGlySerCysGlu----- 884
||| ||| |||||
QY 883 AGGGTGCACACTGCAGGAGGTTCAGGTGCTCCCTGAAGAAATTACTGCGAGTGTATAGGCCCA 942
||| ||| |||||
Db 885 -----TrpCysArgCysGlnAlaGlyG1 892
||| ||| |||||
QY 943 AATTATGTGT----- 952
||| ||| |||||
Db 892 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 912
||| ||| |||||

```
QY 953 -TCCTCTATTTCG-----AAATGCATTGGTTGCAAAATATAT---GAAGAAGCCAGAACG 1005
Db 912 gglySerCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 930
QY 1006 AAAGACACTAATGACATGCCAATACATACATGCAGACTGGAGGTTTGGAGGCGCCATTA 1065
Db 931 -----GluGlySerArgTrp 935
QY 1066 CTGCGCACCAAGAAATTTTCAGGACTTCCAGATTCAGTCACGATAGCGCGCTTCCTC 1125
Db 935 pValProPro-----AspSerAlaCysSerSe 944
QY 1126 ATGCATCTCTGGGAGGTGGTGGAGGCCACATGCGCC-----TGCCCTG-----CT 1170
Db 944 rcysValCysHisGlyGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 963
QY 1171 TCCTCAGGAGNAGAGGCGGAGAAAGACACATGCTCCAGTGCCTGGCGAGACGATGAT 1230
Db 963 salGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 977
QY 1231 CTGCGAGGAATTGGGAAGGTGCTTATACAGATTCTCCACACTGAGTTTAAATCTAAGGG 1290
Db 978 -----SerAspCysGluHisGluGI 984
QY 1291 ATGTGAATGGAG-----TAGAGTATAAGTGTGAATGCATGTT 1329
Db 984 yArgLysTrpGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysII 1004
QY 1330 GATTTTGTTAGTCTAGAAATCTAGTTTGAAGAGATGTTAGGGGACATGAGCT 1389
Db 1004 e----- 1004
QY 1390 GCCTCTGAGCAACACAGGCTCCCTGCATCCCTGGGCCAGGAGTTTACTCAGAGC 1449
Db 1005 -----CysGluProGlnProGluGlyProSerLeu----- 1015
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCTTTTCTCAGGA----- 1489
Db 1016 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 1033
QY 1490 -----GGTCAGTGGCTGAGCATTTG----- 1510
Db 1033 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 1053
QY 1511 -----TTCTCTGGCCCA-----GA 1524
Db 1053 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTrpThrCysGlnCysGI 1073
QY 1525 GGAGAGAGCTTGGTTCCCAT-----GTCTGGGAGAGTCTGCGAGGCGGGG 1546
Db 1073 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGI 1093
QY 1547 -----GTCTGGGAGAGTCTGCGAGGCGGGG 1575
Db 1093 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaG 1113
QY 1576 AGGCGCAGACAGGCGCTCGGAGAGCTCACCTGTGTGCTCTCTCTCAGAGATGT- 1634
Db 1113 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 1131
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
Db 1131 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla---- 1149
QY 1660 RACCTAATGGTTCTTGTGTTTTCATATTTTGAATAAGTTCTCCGATGGGC 1719
Db 1150 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 1164
QY 1720 TGTGTGATACCCTAAATCTCTAGAGACTACTGAACACCTAAGATTTTCTGTAGC 1779
Db 1165 CysCysGluArg-----CysGln 1170
QY 1780 GTAGATATTTCCTCA-----GAGACACGCGAACTGTGCTAGTCTTTCTTAAGCCCCC 1830
```

```
Db 1171 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 1185
QY 1831 GGGAGACGACGCAATGGGCTCGCAGGCCAGGCTTGCACACGATCTCTTGAGTTAGA 1890
Db 1186 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetaIa 1201
QY 1891 GGACTTAAATATATCCAGTTTCTTCTGTGTCTTCTACTTGAATTTGGA 1938
Db 1202 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1217

RESULT 8
US-10-189-971-16
: Sequence 16, Application US/10189971
: Publication No. US20030028907A1
: GENERAL INFORMATION:
: APPLICANT: Walke, D. Wade
: APPLICANT: Scoville, John
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucle
: TITLE OF INVENTION: Same
: FILE REFERENCE: LEX-0360-USA
: CURRENT APPLICATION NUMBER: US/10/189,971
: CURRENT FILING DATE: 2002-07-03
: PRIOR APPLICATION NUMBER: US 60/302,949
: PRIOR FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: US 60/315,634
: PRIOR FILING DATE: 2001-08-29
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 16
: LENGTH: 1251
: TYPE: PRT
: ORGANISM: homo sapiens
US-10-189-971-16

Alignment Scores:
Pred. No.: 6.26e-05 Length: 1251
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 9 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-16 (1-1251)
QY 14 GGCGAAGCTCCGGGGCGCAGCGAGCGGGGAGCTCTCGGGAGTACCCCGGAT 73
Db 222 GlyCysAlaPheGlyGlyLysGluTrpProSerGlyAlaAspPheProHisProSerAsp 241
QY 74 CCCAG-----AGCTCAGCGCTGGAGGAGCTCGCGCTCTCTGTCAGG 114
Db 242 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 261
QY 115 CCCCGCAGCCCGCG-----CCTGCAACGTGCACCTCTCTCTCGTGC 159
Db 262 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 279
QY 160 TACCGCGCACCGCAGCCCGGGTGTTCCTCCCTGGCGG----- 201
Db 280 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 299
QY 202 -----CCTGGGTCTCGAAGAGGCTCCCGCGGGGCTCCGATGATC 246
Db 300 GluTrpPheSerProProGlyAsp-----ProCysArgArgCysLeu 313
QY 247 CCAGTTGAATCAAGTAAGCAGGTGTACTACTACAAGTAATAATCCGGAAGAAC 306
Db 314 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 325
QY 307 TTTCAGATCTCTTCTGCTCAGGAATCCTGTTCGAAGTTCCTGCTCCAGGACTAGA 366
Db 326 CysProProAlaProCysAlaHis----ProArgGlnGlyProCysCysProSerCysAsp 344
```


QY 367 GGATGCTCTCTG-----CTGTCTCTTAAGAAAGATC---CAACCCAAATGT 411
|||||
Db 345 GlyCysLeuTyrglnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL 364
|||||
QY 412 GATATGCCAAATGAAAGGGGGCACAAATGCTATGATATGACAAATCTAGACAAATCTAGACAAAGAGA 471
|||||
Db 364 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysG1 379
|||||
QY 472 ACTAAAGCAGCTC-----CAATTGGTCTCT----- 496
|||||
Db 379 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 399
|||||
QY 497 -----CAGTATCAAGATCAAAATAATATCTACAGTCAGATGCTCCCTAAACC 543
|||||
Db 399 pCysAspGlyCysGluTyrglnGlyLeuSerTyrglnGluSerAsnGlnGluPheProAspPr 419
|||||
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAATTT 588
|||||
Db 419 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 435
|||||
QY 589 AAATCTCATTACACAACTTGAGGAGCCCTTACCATCGGTAGTCAACGGGTCTCTCTTT 648
|||||
Db 436 -----GlyArgProCysGluProProGlyCysSerH1 447
|||||
QY 649 CCCC----- 652
|||||
Db 447 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrglnGlyVa 467
|||||
QY 653 -----TGGGATCAACTCTTCCAGGACCA-----CCAAATAACTTTGGCTGG 696
|||||
Db 467 lThrThrAlaSerGlyGluTyrglnGluProAspProLeuAspProThrCysSerLeu----- 485
|||||
QY 697 GTACTGTGACTGCTTTGCCAGTGGGACTTTTGCACACAC----- 736
|||||
Db 486 ----CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 504
|||||
QY 737 -----TGCAATTGTAATAATTTGTTGTCACCAACTTTGCATCA 771
|||||
Db 504 uCysProHisProSerProGlyProCysPheCysProValCys----- 518
|||||
QY 772 TGATATTGAACGGTTTAAAGCCATTAAAGCATGCTCTGGTAGAAATCCAGAGCTTTCCA 831
|||||
Db 519 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 530
|||||
QY 832 GCCAAAATTTGGAGGGCCAAATTTGGCAATGTCAGAGCCCGACACACAAAGGGTGCA 891
|||||
Db 530 pGlyGluGluPheGlyGlyProAlaGlySerCysGlu----- 542
|||||
QY 892 CTGAGGAGGTGAGGTGCTGCTGAAGAATTTACTGCGAGTGTCTATGAGGCCCAATTTATGTG 951
|||||
Db 543 -----TrpCysArgCysGlnAlaGlyValSerCy 553
|||||
QY 952 T-----TCCTCTAT 960
|||||
Db 553 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCy 573
|||||
QY 961 TTGC-----AAATGCTATTGCAAAATTTAT-----GAAGAAGCCCGACAGAAAGACT 1014
|||||
Db 573 sCysProArgCysArgGlyCysLeuAlaHisGlyGluHisPro----- 588
|||||
QY 1015 AATGAGCATGCCAAACTACATGCGAGTGGAGGTTTGGAGGCGACATTACCTGCCAC 1074
|||||
Db 589 -----GluGlySerArgTrpValProPr 596
|||||
QY 1075 AACGAAATTTTCAGGACTTCCAAGATTACGACATAGCGGCCCTTCCATCATCATCTC 1134
|||||
Db 596 O-----AspSerAlaCysSerSerCysValCy 605
|||||
QY 1135 CTGGGAGGTGGTGGAGGCCACATGCGCC-----TCGCTG-----CTGCTCAGGG 1179
|||||
Db 605 sHisGluGlyVal-----ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 624

QY 1180 AGAAGAGCCGAGAGAAACACTGCTCCAGTGCCTGGCAGCAGATGATCTCTGGAGGA 1239
|||||
Db 624 oArgGlnGlyProHisAspCysCysProGlnCys----- 635
|||||
QY 1240 ATTGGAAGGTGCTTATCAGACATCTCCACACTGAGTTTAAATCTAAGGATTTGAAAT 1299
|||||
Db 636 -----SerAspCysGluHisGluGlyArgLysTy 645
|||||
QY 1300 GGAG-----TAGAGTATAAGTGTGAATGATCTTTGATTTTGTCT 1338
|||||
Db 645 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 662
|||||
QY 1339 TTAGTCTAGAAATCTCTAGTTTAAAGAGGATGTTTAAAGGCAACATGAGGCTGCTCTGCA 1398
|||||
Db 663 -----CysG1 664
|||||
QY 1399 GCACAAACAGGCTCCCTGCAATCCCTGGGCCAGGAGTTTACTCAGAGCTCTCTGAAG 1458
|||||
Db 664 uProGlnProGluGlyProProSerLeu-----Ar 674
|||||
QY 1459 ATGT-----GGCAACCCATGCCCTTTTCTGAGGA-----GGTGC 1494
|||||
Db 674 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 694
|||||
QY 1495 ATGGCCTGAGCAATTGT----- 1510
|||||
Db 694 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 714
|||||
QY 1511 -----TTGCTCTGGCCCA-----GAGGAGAGAGC 1533
|||||
Db 714 uLeuGlySerGluLeuAlaProProAspProCysTyrglnCysGlnCysGlnAspLeuTh 734
|||||
QY 1534 TTGGGTTCACATA----- 1546
|||||
Db 734 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 754
|||||
QY 1547 -----GTCTCTGGAGAGTGTCTGCGAGGGCGGGGAGGAGAG 1584
|||||
Db 754 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArg 774
|||||
QY 1585 CAGGCCCTGGGAGAGCTACTCTGTGTCAGTCTTCTCTCAGAGATGT----- 1634
|||||
Db 774 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 792
|||||
QY 1635 --TGCTCTGGAGC-----TGCTCTGATGAAACCCCTAAT 1668
|||||
Db 792 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 807
|||||
QY 1669 GGTTCCTGTTGTTGTTTCAAAATTTAGAAAATAAGTTCTCCGGATGGGCTGTGTGTAT 1728
|||||
Db 808 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-SerCysCysGlu 825
|||||
QY 1729 ACCATTAAATCTCTAGAGACTACTGAACACTTAAGATTTCTGTACGCTAGATATT 1788
|||||
Db 826 Arg-----CysGlnAlaProThr 831
|||||
QY 1789 TCCCCA-----GAGACACCGCAACTGCTCAGTCTTCTTAAGCCCCCGGAGAGCG 1839
|||||
Db 832 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 846
|||||
QY 1840 AGCAATGGGCCCTCCAGGCCAGGCTTGCAACAGCATGCTCTGAGTTAGAGGACTTAA 1899
|||||
Db 847 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 862
|||||
QY 1900 ATTATCCAGTTTCTCTGTGTTTCTACTTGAATTTGGA 1938
|||||
Db 863 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 875
|||||

RESULT 9

US-10-189-971-24

; Sequence 24, Application US/10189971

; Publication No. US20030028907A1

; GENERAL INFORMATION:

APPLICANT: Walke, D. Wade
APPLICANT: Scoville, John
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907alel Human Kielin-like Proteins and Polynucleotides
TITLE OF INVENTION: Same
FILE REFERENCE: LEX-0360-USA
CURRENT FILING DATE: 2002-07-03
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302,949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315,634
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 1342
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-24

Alignment Scores:
Pred. No.: 6.42e-05 Length: 1342
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 9 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-24 (1-1342)

Qy 14 GCGAAGCTCGCGGCGGACAGCGGCGGAGTCTCTCGGGAGTACCCGGAT 73
Db 313 GlyCysAlaPheGlyGlySerProSerGlyAlaAspPheProHisProSerAsp 332
Qy 74 CCCAG-----AGTCACGCGCTGGAGGAGTCCGCGTCTCTCGAGG 114
Db 333 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 352
Qy 115 CCCGCGACGCGCG-----CCTCAAGTGCACCTCTCTCGCTCGGTGC 159
Db 353 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 370
Qy 160 TACCGGCGACCGCGCGCGGTGTTCCTCCCTCGGGCG----- 201
Db 371 ProAlaAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 390
Qy 202 -----CCTGGTCTCGCGAAGGAGCTCCACCGCGGCGTCCGATGATC 246
Db 391 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 404
Qy 247 CCAGTTGAATCAAGGTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGCAAC 306
Db 405 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 416
Qy 307 TTGCAGAACTCTTGTCTCAGGAATCCTGTTCAGAGTCCCATGTCGCCAGAACTAGA 366
Db 417 CysProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 435
Qy 367 GGATGCTCTCG-----CTGTCTCTTAAGAAAGATTC---CAACCCCAATGGT 411
Db 436 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL 455
Qy 412 GATATGCCAAATTGAAGGGGCGACACAAATCTATGTATAGCAATCTCAGAACAGAGA 471
Db 455 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 470
Qy 472 ACTAAAGCACTC-----CATTTGGTTCCT----- 496
Db 470 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 490
Qy 497 -----CAGTATCAAGTCAAAATAATTAATCTACAGTCAGATGTCCTAAACC 543
Db 490 pCysAspGlyCysGluTyrLeuGlyGlySerTyrLeuSerAsnGlnGluPheProAspPr 510

Qy 544 A-----ATGACTGCTTTAGTAGGAGATTTTTCAGCAGCATCAACAAAAT 588
Db 510 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 526
Qy 589 AAATCTCATTAACAACAACCTTACAGGAGCCTTACCATCGGTAGTCAACGGGTGCTTT 648
Db 527 -----GlyArgProCysGluProGlyCysSerHis 538
Qy 649 CCCC----- 652
Db 538 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 558
Qy 653 -----TCGGGATCAACTTCCAGGACCA-----CCAAAAATAACTTTGGCTGG 696
Db 558 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 576
Qy 697 GTACTGTGACTGCTTTGCCAGTGGGACCTTTTGCACAAC----- 736
Db 577 ---CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 595
Qy 737 -----TGCAATTGTAATAATTTGTCACAACTTGCATCA 771
Db 595 uCysProHisProSerProGlyProCysPheCysProValCys----- 609
Qy 772 TGATATTGAACGGTTAAAGCCATTAAAGGCATGCTTGGTAGAAATCCAGAAAGCTTTCCA 831
Db 610 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 621
Qy 832 GCCAAAAATGGGAAGGCCAAATTTGGCAATGTCAAGCCCAAGCACAACAAAGGTGCAA 891
Db 621 pGlyGluGluPheGluGlyProAlaGlySerCysGlu----- 633
Qy 892 CTGAGGAGGTGAGCTGCCTGGAAGATTTACTGCGAGTGTCTAGGCCCAAAATATGTG 951
Db 634 -----TrpCysArgCysGlnAlaGlyGlnValSerCys 644
Qy 952 T-----TCTTCTAT 960
Db 644 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCys 664
Qy 961 TTGC---AATGCAATGTTGCAAAAATAT---GAGAAAGCCCAAGCAACAAAGACT 1014
Db 664 sCysProArgCysArgGlyCysLeuAlaHisGlyGluHisPro----- 679
Qy 1015 AATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGCGAGCCATTACCTGCCACC 1074
Db 680 -----GluGlySerArgTrpValPro 687
Qy 1075 AACGAAATTTTCAGGACTTCCAAGATTACGTACAGTAGGCGCTTCTCATCATCTC 1134
Db 687 O-----AspSerAlaCysSerSerCysValCys 696
Qy 1135 CTGGAGGTGTTGAGGCCACATCGGC-----TGCTG-----CTTGTCTCAGG 1179
Db 696 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 715
Qy 1180 AGAAGGCGGAGAAAGAACACTCTCAAGTGCCTGCAGAGCAGATGATCCTGGAGGA 1239
Db 715 oArgGlnGlyProHisAspCysCysProGlnCys----- 726
Qy 1240 ATTTGGAGGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGATTTGAAAT 1299
Db 727 -----SerAspCysGluHisGluGlyArgLys 736
Qy 1300 GGAG-----TAGATATAAGTGAATGCAATGCTGTTGATTTGCT 1338
Db 736 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 753
Qy 1339 TTAGTCTAGAAATCTCTAGTTTAGAAAGGATTTTAGGGGAACATGAGGCTGCTCTCA 1398
Db 754 -----CysGln 755

```
QY 1399 GCAACAAACGAGCTCCCTGCAATCCCTGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAG 1458
Db 755 uProGlnProGluGlyProProSerLeu 1458
QY 1459 ATGT---GGCAACCCATGCCCCCTTTCTGAGGA-----GGTGC 1494
Db 765 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 785
QY 1495 ATGGCTCTGAGCATGT--- 1510
Db 785 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 805
QY 1511 -----TTGCTGGGCCA-----GAGGAGAGAGC 1533
Db 805 uLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGlnAspLeuTh 825
QY 1534 TTGGGTTCACATA----- 1546
Db 825 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 845
QY 1547 -----GTCTGGGAGAGTGTCTGCAGGCGCGGAGGAGGAGCAG 1584
Db 845 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 865
QY 1585 CAGGCCCTGGGAGAGCTCACTCTGTGCTGACTCTTCTCTCAGAGAAATGT----- 1634
Db 865 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCysT 883
QY 1635 --TGCTCTGGAGC-----TGCTGTGATGAACACCTAAT 1668
Db 883 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 898
QY 1669 GGTTCCTGTTCTTTTCAAAATATTAGAAATAAGTTCTCCGGGATGGCTCTGTCTGAT 1728
Db 899 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-SerCysCysGlu 916
QY 1729 ACCATTAAATCTCTAGAACTACTGAACACCTAAAGATTTCGTGACCTAGATATT 1788
Db 917 Arg-----CysGlnAlaProThr 922
QY 1789 TCCCCA-----GAGACACGCACTGTCAGTCTTCTTAAGCCCGCGGAGAGCG 1839
Db 923 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 937
QY 1840 AGCAATGGGCTCCGAGCCAGGCTTGACCCAGCATGCTTGTAGCTAGAGCACTAAA 1899
Db 938 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetalGlyThrVal 953
QY 1900 ATTATCCAGTTCTCTGTGTTCTACTTGAATGTGGA 1938
Db 954 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 966
```

RESULT 10

```
US-10-189-971-10
; Sequence 10, Application US/10189971
; Publication NO. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20030028907A1el Human Kiellin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR APPLICATION DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1512
```

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-10

Alignment Scores:      6.69e-05      Length:      1512
Pred. No.:            157.50          Matches:     172
Percent Similarity:    27.78%          Conservative: 60
Best Local Similarity: 20.60%          Mismatches:  231
Query Match:           4.12%          Indels:      373
DB:                     9              Gaps:        47

US-09-743-237-3 (1-2134) x US-10-189-971-10 (1-1512)
QY 14 GCGAAGCTCGGGGGGGGCGACAGCGCGGGAGCTCTCGGGGAGTACCCGGGAT 73
Db 483 GlyCysAlaPheGlyGlyGlyProSerGlyAlaAspPheProHisProSerAsp 502
QY 74 CCAG-----AGCTCAGCGCGCTGGAGGACGTCGCGCTCCTGCAGG 114
Db 503 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 522
QY 115 CCGCGAGCGCGCCG-----CCTGCAACGTGACACTCTCTCTCTCGCTGC 159
Db 523 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 540
QY 160 TACCCCGGACCGCGCGCGGTCTTTTCCCTCGGGCG----- 201
Db 541 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 560
QY 202 -----CCTGGTCTCGAAGAGCGCTCCACCGCGCGCTCCGCATGATC 246
Db 561 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 574
QY 247 CCAGTTGAATCAAGTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGAAC 306
Db 575 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 586
QY 307 TTGTCAGAATCTCTTCTCAGGAATCCTCTTCAAGTTCCTCCATGGTCCCGAGACTAGA 366
Db 587 CysProProAlaProCysAlaHis-----ProArgGlnGlyProCysCysProSerCysAsp 605
QY 367 GGATGCTCTCTG-----CTGTCTCTTAAGAAGATTC---CAACCCCAATGGT 411
Db 606 GlyCysLeuTyrGlnGlyGlyPheAlaSerGlyGluArgPheProSerProThr-Al 625
QY 412 GATATCCCAATTGAAGGGGCGACAAATGCTATGATAGACAATCTATAGACAAGAGA 471
Db 625 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGlu 640
QY 472 ACTAAAGCACTC-----CATTTGTTCTCT----- 496
Db 640 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 660
QY 497 -----CAGTATCAAGATCAAAATAATATTACTACAGTCAGATCTCCCTAAACC 543
Db 660 pCysAspGlyCysGlyTyrLeuGlyGlySerTyrLeuSerAsnGlnGluPheProAspPr 680
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAAT 588
Db 680 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 696
QY 589 AAATCTCATTACACAACACTTTCAGGGAGCCCTTACCATCGGTAGTCAACGGTCTCTTT 648
Db 697 -----GlyArgArgProCysGluProProGlyCysSerHl 708
QY 649 CCCC----- 652
Db 708 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 728
QY 653 -----TCGGGATCAACTCTTCCAGGACCA-----CCAAATAAATTTGGCTGG 696
Db 728 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 746
```

```

Qy 1534 TTGGGTTCCCAATA-----154
Db    |||::: |||
Qy 995 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHISTh 1015
Db    |---GTCTGGAGACTGTCGCAGCGCGCGGAGGCCACAG 1584
Qy 1547 -----||| |||:::~||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1035
Qy 1585 CAGGCCCTCGGGAGAGCTCAC'TGTGTCAGTCTCTCTCTCAGAGAATG-----1634
Db 1035 rg--ValAlaAspGlyGluSerTrpArgASP---ProSerAsnAlaCysIleAlaCyst 1053
Qy 1635 --TGCTCTGGAGGC-----TGCTCTGCATTGAACCCTTAAT 1668
Db 1053 hrCysHisArGgLyHisValGlucYsHisLenLuglucYsGlnAla-----1068
Qy 1669 GGYTTCTGTGTTCTTTTTCAAATATTATTTAGAANAATGTTCCCGGATGGCGTGTGTGAT 1728
Db 1069 ----LeuSerCysProHisGlyTrpAlaLaLYSValPrOGlnAlaAsp-SerCysCysGLU 1086
Qy 1729 ACCATTAAAAATCTCTAGAACTACTGAACACCTAAAAGATTTCCTGTAGCGTAGATATT 1788
Db 1087 Arg-----CysGlnAlaProThr 1092
Qy 1789 TCCCCA-----GAGACACGCAACTGTGACTGTTCTCTAAGGCCCGGGAGACGC 1839
Db 1093 GlNserCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1107
Qy 1840 AGCATGGGCTCCAGCGCAGCGTTCACCAACATGTC'TTGAGTATAGAGGACIT'AAA 1899
Db 1108 ---TriPhrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1123
Qy 1900 ATTATCCAGTTCTCTGTGTTCTTACTTGAATTGTGGA 1938
Db 1124 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1136

RESULT 11
US-10-189-971-12
; Sequence 12, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like proteins and
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1570
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-12

Alignment Scores:
Pred. No.:      6,77e-05      Length:      1570
Score:          157.50       Matches:     172
Percent Similarity: 27.78%   Conservative: 60
Best Local Similarity: 20.60% Mismatches:    231
Query Match:        4.12%    Indels:         373
DB:                ~~~~~    Gaps:           47

US-09-743-237-3 (1-2134) x US-10-189-971-12 (1-1570)

Qy 14 GGCGAAGCTCGCGGGGGGCGACAGCGACGGCGGGGAGTCTCTCGGGGAGTACCCCGGAT 73
Db  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 541 GlyCysAlaPheGlyGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 560
QY 74 CCAG-----AGCTCAGCGCGCTGGAGGACGTCGCGCTCCCTCGCAGG 114
Db 561 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 580
QY 115 CCCCGAGCGCGCG-----CTGCAAGTGCACCTCTCTGCTCGCTGC 159
Db 581 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 598
QY 160 TACCGCGCGCAGCGCGCGGGTGTTCGCCCTGGCGG----- 201
Db 599 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 618
QY 202 -----CCTGGTCTCGCGAAGGAGCTCCACCGCGGCTCGCATGATC 246
Db 619 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 632
QY 247 CCAGTTGAATCAAGTAAAGAGGTGGTACTACTACAAGTAATAATCCGGAAGAAGCAAC 306
Db 633 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 644
QY 307 TTTCGAGAATCTTCTGCTCAGCAATCTGTTGCAAGTTCCTCATGTCGCCAGCACTAGA 366
Db 645 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 663
QY 367 GGATGCTCTCTG-----CTGTCTCTTAAGAAAGATTC---CAACCCAAATGCT 411
Db 664 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 683
QY 412 GATATGCCAAATTAAGAGGGGCGCACAAATGCTATGATATAGACAATTTCTAGAACAGAGA 471
Db 683 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 698
QY 472 ACTAAAGCAGCTC-----CATTTGGTTCCT----- 496
Db 698 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 718
QY 497 -----CAGTATCAAGATCAAAATAATATCTACAGTCAGATGTCCTCAAAACC 543
Db 718 pCysAspGlyCysGluTyrLeuGlyGlySerTyrLeuSerAsnGlnGluPheProAspPr 738
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTGGCAGCATCAACAAAAT 588
Db 738 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 754
QY 589 AATCTCATTACACACACTTGAGGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 755 -----GlyArgArgProCysGluProProGlyCysSerHi 766
QY 649 CCCC----- 652
Db 766 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 786
QY 653 -----TCGGGATCAACTTTCCAGGACCA-----CCAAAATAACTTTGGCTGG 696
Db 786 lPThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 804
QY 697 GTACTGTGACTCTTTCGCGAGTGGGACTTTTGGCAACAC----- 736
Db 805 -----CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 823
QY 737 -----TGCAATTTGTAATAATTTGTTGCAACAACTTGCATCA 771
Db 823 uCysProHisProSerProGlyProCysPheCysProValCys----- 837
QY 772 TGATATTGAACGGTTTAAAGCATTTAAGCATGCTTGGTAGAATTCAGAACTTTTCCA 831
Db 838 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 849
QY 832 GCCAAAATTTGGGAAGGCCAATTCGGCAATCTCAAGCCCCAGCAACAAAGGTGCA 891
Db 849 pGlyGluGluPheGluGlyProAlaGlySerCysGlu----- 861

QY 892 CTGCAGGAGGTGAGGCTGCCTGAAGAATTACTCGAGTGTCTATGAGGCCCAAAATTATGTG 951
Db 862 -----TrpCysArgCysGlnAlaGlyGlnValSerCy 872
QY 952 T-----TCTTCTAT 960
Db 872 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCy 892
QY 961 TTGC---AAATGCATTGTTGCAAAATAT---GAAGAAACCCAGACGAAAGACACT 1014
Db 892 sCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 907
QY 1015 AATGAGATGCCAAACTACATCAGACTGGAGGTTTGAAGGAGGCACATTACTTGCACCC 1074
Db 908 -----GluGlySerArgTrpValProPr 915
QY 1075 AACGAAATTTTCAGGACTTCCAAGATTACGTACCATAGGCGGCTTCTCTCATCATCTC 1134
Db 915 O-----AspSerAlaCysSerSerCysValCy 924
QY 1135 CTGGGAGGTGGTGGAGGCCACATGCGCC-----TGCCTG-----CTTGTCTCAGGG 1179
Db 924 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 943
QY 1180 AGAAGAGCGCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAGGA 1239
Db 943 oArgGlnGlyProHisAspCysCysProGlnCys----- 954
QY 1240 ATTTGGAAGGTCTTATCACAGATTCTCACACTGATGTTAAATCTAAGGATTGAAAT 1299
Db 955 -----SerAspCysGluHisGluGlyArgLysTy 964
QY 1300 GGAG-----TAGAGTATAAGTGTGAATGCATGCTTGTGATTTGTGTC 1338
Db 964 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 981
QY 1339 TTAGTCTAGAAATCTCTAGTTTAGAAGGATGTTTAGGGGAACATGAGCTGCTCTGCA 1398
Db 982 -----CysGln 983
QY 1399 GCACAACACCGAGCTCCCTGCTCATCCCTGGGCCAGGAGGATTTACTCAGAGCTCTCTGAAG 1458
Db 983 uProGlnProGluGlyProProSerLeu-----Ar 993
QY 1459 ATGT---GGCAACCCATGCCCTTTCTTGAGGA-----GGTGC 1494
Db 993 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 1013
QY 1495 ATGGCTGTGACATTCGT----- 1510
Db 1013 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 1033
QY 1511 -----TTGTCTGCGCCA-----GAGGAGAGAGC 1533
Db 1033 uLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGlnAspLeuTh 1053
QY 1534 TTGGTTCCCAT----- 1546
Db 1053 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1073
QY 1547 -----GTCTCTGGAGAGTGTCTGCGAGGCGCGGAGGAGGAGAG 1584
Db 1073 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1093
QY 1585 CAGGCCCTCGGAGAGCTCCTCTGTCGACTCTTCTCTCTCAGAGAATGT----- 1634
Db 1093 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 1111
QY 1635 --TGCTCTGGAGGC-----TGCTCTGCATGAAACCCCTAAT 1668
Db 1111 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 1126

QY 1075 AACGAAATTTTCAGGACTTCCAGATTCCAGATAGCGGCTTCCTCATGCATC 1134
Db 973 O-----AspSerAlaCysSerSerCysValCy 982
QY 1135 CTGGAGGTGGTGGAGGCCACATGCGCC-----TGCCRG-----CTTGCTCAGGG 1179
Db 982 shIsGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 1001
QY 1180 AGAAGAGCCGAGAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATCATCTGGAGGA 1239
Db 1001 oArgGlnGlyProHisAspCysProGlnCys----- 1012
QY 1240 ATTGGAAGGTCTTATCACAGATTCTCCACACTGAGTTTAATCTAAGGGATTGAAAT 1299
Db 1013 -----SerAspCysGluHisGluGlyArgIlyTy 1022
QY 1300 GGAG-----TAGAGTATAAGTGTGAATGCATGTTGATTTGTC 1338
Db 1022 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 1039
QY 1339 TTAGTCTAGAAATCTAGTTTAGAAGGATGTTTAGGGGAACATGAGCTGGCTCTGCA 1398
Db 1040 -----CysG1 1041
QY 1399 GCAACAACGAGCTCCCTGTCATCCCTGGCCAGGAGTCTTACTCAGAGTCTCTGAAG 1458
Db 1041 uProGlnProGluGlyProProSerLeu-----Ar 1051
QY 1459 ATGT---GGCAACCATGCCCCCTTTTCTGAGGA-----GGTGC 1494
Db 1051 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 1071
QY 1495 ATGGCTGAGCATGCT----- 1510
Db 1071 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 1091
QY 1511 -----TTGCTTGGGCCA-----GAGGAGAGAGC 1533
Db 1091 uLeuGlySerGluLeuAlaProAspProCysTyThrCysGlnCysGlnAspLeuth 1111
QY 1534 TTGGTTCCCAT----- 1546
Db 1111 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1131
QY 1547 -----GTCTGGGAGAGTGTCTGCAGGGCGGCGAGGAGCAGAG 1584
Db 1131 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1151
QY 1585 CAGGCCCTGCGGAGAGCTCACTCTGCTGCACTCTTCTCTCAGAGAATGT----- 1634
Db 1151 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 1169
QY 1635 --TGCTCTGGAGC-----TGCTCTGCATGAAACCCCTAAT 1668
Db 1169 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 1184
QY 1669 GCTTTCTGTTGTTTTCAAATTTATTAGAAATAGTTCTCCGGATGGCGTGTGTGAT 1728
Db 1185 -----LeuSerCysProHisGlyTrpAlaIlyValProGlnAlaAsp-SerCysCysGlu 1202
QY 1729 ACCACTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCCTGTAGCGCTAGATAT 1788
Db 1203 Arg-----CysGlnAlaProThr 1208
QY 1789 TCCCCA-----GAGACACGCAACTGTCAGTCTTTCCTAAAGCCCCCGGAGAGCGC 1839
Db 1209 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1223
QY 1840 AGGCAATGGGGCTCGCAGGCCAGGCTTCACACAGCATGCTCTGTAGTTAGAGGACTTAA 1899
Db 1224 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1239

QY 1900 ATTATCCAGTTTCTCTGTGTTTCTTACTTGAATTTGTGA 1938
Db 1240 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1252
RESULT 13
US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; APPLICANT: Scott, Matthew
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 60/311,720
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-213-509-5
Alignment Scores:
Pred. No.: 0.00552 Length: 4123
Score: 137.00 Matches: 153
Percent Similarity: 27.03% Conservative: 50
Best Local Similarity: 20.37% Mismatches: 208
Query Match: 3.58% Indels: 340
DB: Gaps: 44
US-09-743-237-3 (1-2134) x US-10-213-509-5 (1-4123)
QY 7 GGGTCAAGCGAAGCTCGCGGGG-----GCGACAGCAGCGCGGGGAGC 51
Db 2820 GlyAspArgGlnGluLeuGlnGlyCysHisThrValCysGlyThrGlyIleAlaGlySer 2839
QY 52 TCCTCGGGAGTACCCCGGATCCAGACTCA----- 84
Db 2840 LeuGlyAlaGlyValProProSerSerSerGlnPheCysThrLeuArgThrHisGlyMet 2859
QY 85 -----GCGCGC 90
Db 2860 GlyProThrAspHisSerThrTrpGlyIleGluValPheGlyTrpThrProThrSer 2879
QY 91 TGGAGAGC---TCGCGCTCTGCG-----AGGCCCGC 120
Db 2880 TrpSerSerCysSerGlnSerCysLeuAlaProGlyGlyGlyProGlyTrpArgSerArg 2899
QY 121 AGCCGC----- 126
Db 2900 SerArgLeuCysProSerProGlyAspSerSerCysProGlyAspAlaThrGlnGlu 2919
QY 127 -----CCGCTCGAACGTGCACT-----TCCTGTCTCT--- 153
Db 2920 ProCysSerProProIleGluCysThrGlyPheCysAlaProGlyCysThrCysProPro 2939
QY 154 -----CGCTGCTACCCCGCACCGCCCGCGGGTGT----- 186
Db 2940 GlyLeuPheLeuHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeu 2958
QY 187 -----TTTGCCCTGGGCGCTGGG-----TCCTGCCAAGGAGCTCC 225
Db 2959 HisGlyGlnLeuTyAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysThr 2978
QY 226 CACCCGGGCGTCCGATG-----ATCCAGTTCAAATCAAGGTAAGC 267
Db 2979 CysValSerGlyIlycMetAlaCysThrSerGluArgCysProValAlaCysGlyTrpSer 2998
QY 268 AGGTGGTACTACTAAGTAATAATCCGGAAGAACAACTTTGCAGAAATCTTCTTGCTCA 327
Db 268 AGGTGGTACTACTAAGTAATAATCCGGAAGAACAACTTTGCAGAAATCTTCTTGCTCA 327

Db 2999 ProTrpThrLeuTrp-----SerLeuCysSer 3007
Qy 328 GGAATCCTGT----- 337
Db 3008 Cys-SerCysAsnValGlyIleArgArgArgPheArgAlaGlyThrAlaProProAlaAl 3027
Qy 338 -----TGCAGTTTCCCATGGTCCAGGAACACTAGAGATGCTCTCTGCTG 381
Db 3027 aPheGlyAlaGluCysGlnGlyPro-----ThrMetGluAlaGluPheCy 3043
Qy 382 TTCCTTTAAGAAAGATTCCCAACCAATGGTGATATGCAATTTGAAAGGGGCGCACACAAT 441
Db 3043 sSerLeuArgProCysProGlyProValProGlyMetCysProArgAspLysGlnTrpLe 3063
Qy 442 GCTATGATAGACAAT---TCTAGAACAGAGAACTAAAGCACTCCATTGGTTGCTCTCA 498
Db 3063 uAspCysAlaGlnGlyProAlaSerCysAlaGluLeuSerAla-----ProAr 3079
Qy 499 GTATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTTAACCAATGACTGCTTTAGT 558
Db 3079 gGlyThrAsnGlnThrCysHisProGlyCysHisCysProSerGlyMetLeuLeuLeuVa 3099
Qy 559 AGGG---AGATTTTGGCAGCATCAACAAATAATCTATTACACAACTTGAGGG 615
Db 3099 lSerProArgGlyHisProGlyProLeuGlyAlaSerVal-----GlnProProVa 3116
Qy 616 AGCCTTACCAACGCTAGTCAACGGGTCTGCTTTCCCTCGGGATCA----- 661
Db 3116 lAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyTrpGlyProTr 3136
Qy 662 -----AC 663
Db 3136 pGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArgThrArgAl 3156
Qy 664 TCITCCAGGACCACAAATAACTTTGGTGGTACTGTGAC---TGCITTCACAGTGG 720
Db 3156 aCysAspGlnProProGlnGlyLeuGlyAspTrpCysGlnGlyProArgAlaGlnGl 3176
Qy 721 GGACTTTTTC-----AACAACTGCAAT----- 742
Db 3176 yGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGluGlyAlaGl 3196
Qy 743 -----TGTAAATAATTTGTGCAACAACATGTCATCATGATATGA 780
Db 3196 uTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis----- 3213
Qy 781 ACGTTTAAAGCCATTAAAGCATCTCTGGTAGA---AATCCAGAACTTTCAGCCAAA 837
Db 3214 -----CysValTrpArgCysGlnProGlyCysTrpCysPro-- 3225
Qy 838 AATTGGGAAGGCCAATTTGGCAAT-----GTCAAGCCCCAGCACAA 879
Db 3226 -----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHis-- 3242
Qy 880 CAAAGGTGCAACTGC----- 895
Db 3243 -----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLe 3260
Qy 896 ---AGAGGTGCGCTGAGAGATTAATCTGCGAGCTGATGAGCCCAAAATTATGTG 951
Db 3260 uAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnC 3278
Qy 952 TTCCTCTATTGCAATGCAATTTGTTGTCAAAATAATTATGAAGAACGCCAGAACGAAAGAC 1011
Db 3278 sThrAspLeu---ProCysProAspCys----- 3286
Qy 1012 ACTAATGAGCATGCCAAACTACATGCACTGGAGGTTTGAAGGAGCCCATTTACCTGCC 1071
Db 3287 -----GlyGlyGlnSerLeuHis----- 3293
Qy 1072 ACCAACGAAATTTTCAGGACTTCCAAAGATTCAGTCAGTACAGTGGGCTCTCCCTCA---TG 1128
Db 3294 -ProCysGly---GlnProCysProArgSerCysGlnAspLeuSerProGlySerValCy 3312

Qy 1129 CATCTCTCTGGAGGTG---GTGGAGGCCACATCGCGCTGCTGCTGCTCAGGAGAAGA 1185
Db 3312 sGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln----- 3329
Qy 1186 GCGCGAGAAAGAACACTGCTCCCAAGTCCCTGGCAGACAGATGATCTGCGAGGAATTTGG 1245
Db 3330 -----LeuSerGlnAspGl 3334
Qy 1246 AAGTGCTTATACAGATTCTCCACACTGAGTTTAAATCTAAGGGATTGAAATGGAGTA 1305
Db 3334 yLeuCysValPro----- 3338
Qy 1306 GAGTATAAAGTGAATGTCATGTTGATTTTCTTCTAGTCTAGAAATCTCTAGTTTAGAAA 1365
Db 3339 -ProAlaHisCysArgCys----- 3344
Qy 1366 GGATGTTTGGGGAACATGAGGTGGCTCTGCAGCAACAACAGGCTCCCTGCA---TC 1422
Db 3345 -----GlnTyrGlnProGlyAlaMetAlaProSe 3354
Qy 1423 CCTGGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAAC-----CCATGCC 1476
Db 3354 rPheValProSerThrCysValAlaGlyIleLeuGlnCysGlnGluValProAspCysPr 3374
Qy 1477 CCTTTTCTGAGGAGTGCATGGCTGAGCATTTGTTCTCTGCCCCAGAGAGAGAGCTTG 1536
Db 3374 o-----AspProGlyValTrpSerSerTrp-----GlyProTrpGluAspCysSe 3389
Qy 1537 GGTTCCTCCATGCTCTGGGAGAGTCTGTCAG----- 1567
Db 3389 rValSerCysGlyGlyGlyGluGlnLeuArgSerArgCysAlaArgProProCysPr 3409
Qy 1568 -GCGGGGAGGAGCAGAGCAGGCGCTCGGAGAGCTCACTCTGTCGACTCTTCTCTCTCA 1626
Db 3409 oGlyProAlaArgGlnSerArgThrCys-----SerThrGl 3421
Qy 1627 GAGAATTTCTCTCTGGAGGCTGCTCTGCA 1655
Db 3421 n-ValCysArgGluAlaGlyCysProAla 3430

RESULT 14

US-10-037-182-36
; Sequence 36, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thybøll, Jili
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2743
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-037-182-36

Alignment Scores:
Pred. No.: 0.0174 Length: 2743
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
DB: 9 Gaps: 28

US-09-743-237-3 (1-2134) x US-10-037-182-36 (1-2743)

```
Qy 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
Db 83 LysLeuValGlyGlyProValAlaGlyAspProAsnGlnThrIleArgGlyGlnTyr 102
Qy 66 CCCGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleasp 122
Qy 96 GACGTC-----CGGCTCTGCAGGCCCGCCGCGCCCTGC 134
Db 123 GlyThrGluArgTrpTrpGlnSerProLeuSerArgGlyLeuGluTyrAsnGluVal 142
Qy 135 AACGTG-----CACTTCCTGCTCGCTGCTACCCGCGCAC 170
Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162
Qy 171 CGCAGCCCGGGGTGTTTGGCCCTGGCGCCCTGGGTCCTGCGAAGGAGCTCCCAACC 230
Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetaspPhe 178
Qy 231 GGGCGTCGCGATGATCCAGTTGAAATCAAGGTAAAGCAGGTGGTACTACTACAAGTAATA 290
Db 179 GlyArgThrTyrGlnPro-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
Qy 291 ATCCGGAAGAAGCAACTTTGCGAATCTTCTGCTCAGGAATCTGTTGCAAGTCCCAT 350
Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
Qy 351 GGTCCCAAGAACTA-----GAGGATGCTCTGCTGCTCTCTTAAGAAA 394
Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaAlaIleCysThrThrGluTyr 219
Qy 395 GATTCACACCAATGGTGATATGCAATGAAGGGGGCACACAATCTATGATATAGAC 454
Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
Qy 455 AATCTAGAACAAAGAACTAAAGCACTCCATTTGGTTCTCAGTATCAAGATCAAAAT 514
Db 236 AsnGlyArgProGlyAlaMet-----AGA 565
Qy 515 AATATCTACAGTCAGATGCCCTAAACCAATGACTGCTTTAGTAGGG-----AGA 565
Db 243 AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg 262
Qy 566 TTTTGGCAGCATCAACAATA---AATCTCATTACACAACACTTGGAGGACCTTA 622
Db 263 PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp-----280
Qy 623 CCATCGGTAGTCAACGGGTCTGCTTCCCTCGGGATCACTCTTCCAGGACCAACAAA 682
Db 281 ProThrValThrArgArgTyrTyr-----SerIleLysasp 293
Qy 683 ATAATTTGGTGGTACTGTACTGCTTTTCCAGTGGGGACTTTTGC-----730
Db 294 IleSerIleGlyArgCysValCysHisGlyHisAlaAspAlaCysAspAlaLysasp 313
Qy 731 -----ACAACCTGC-----AATGCT 745
Db 314 ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys 333
Qy 746 AATAATCTTGCACAACTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGCA---802
Db 334 AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn 351
Qy 803 -----TGCTCTTGCTAGA-----AAT 817
Db 352 SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp 371
Qy 818 CCAGAA-----GCTTCCAGCCAAAATTTGGAGGGCCCAATTTGGGCAAT 862
Db 372 ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGlyGly 391
```

```
Qy 863 GTC-----AAGCCCCAGCACAAAGGGTCAACTGCAGGAGCTAGGCTGCCTG 913
Db 392 ValCysIleAspCysGlnHisThrAlaGlyValAsnCysGluArg-----CysLeu 409
Qy 914 AAGAATTAATCGAGTGTATGAGGCCCAATTTATGTCTTCTTATTTGCAAAATGCATT 973
Db 410 ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys 429
Qy 974 GGTTCAAAAAT-----TATGAAGAAACCCAGAACGAAAGACACTA 1015
Db 430 AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys 447
Qy 1016 ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGGAAAGC 1057
Db 448 TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly 467
Qy 1058 AGCCATTACCTGCCAACCAAGAAATTTTCAGGACTTCCAGATTTCAGTCAGATAGCGG 1117
Db 468 PheProSerCysTyrProThr-----ProSerSerSerAsnAspThrArg 482
Qy 1118 CCTTCTCTCATCATCTCTGGGAGGTGTGGAGGCCACATGGCCTGCTGCTGCTCAG 1177
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln 502
Qy 1178 GGAGAAGAGGCGGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysasp-----Pro 510
Qy 1238 GAATTTGGAAGTGTCTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
Qy 1286 AAGGATTGAATGGAGTAGATAGTATAAAGTGTGAATGC 1324
Db 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543
```

RESULT 15

```
US-10-037-182-2
; Sequence 2, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3695
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-037-182-2
```

Alignment Scores:

Pred. No.:	0.0193	Length:	3695
Score:	130.50	Matches:	111
Percent Similarity:	31.14%	Conservative:	55
Best Local Similarity:	20.83%	Mismatches:	198
Query Match:	3.41%	Indels:	170
DB:	9	Gaps:	28

US-09-743-237-3 (1-2134) x US-10-037-182-2 (1-3695)

```
Qy 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
Db 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
```

Db	83	LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr	102	QY	914	AGAATTACTGCCAGTGCTATGAGGCCAAATATTATGTGTTCTTCTATTTCGCAATGCATT	973
QY	66	CCCGGATC-----CCAGAGCTCAGCGCGCTGGAG	95	Db	410	ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys	429
Db	103	CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp	122	QY	974	GGTTGCAAAAT-----TATGAAGAAAGCCAGAACGAAAGACACTA	1015
QY	96	GACGTC-----GCGTCCTGCAGGCCCGCCGAGCGCCGCGCTGC	134	Db	430	AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys	447
Db	123	GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal	142	QY	1016	ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGGGAAGC	1057
QY	135	AACGTG-----CACTTCCCTGCTCCTCGCTACTACCCGCGCAC	170	Db	448	TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly	467
Db	143	AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValIleLysPheAla	162	QY	1058	AGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCACGTACGATAGGCGG	1117
QY	171	CGCAGCCCGGGGGTGGTGGCCCTGGCGGCTCGAAGAGGACCTCCACACC	230	Db	468	PheProSerCysTyrProThr-----ProSerSerAsnAspThrArg	482
Db	163	AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe	178	QY	1118	CCTTCCTCATGTCATCCTCGGAGGTGGTGAGGCCACATCGCGCTGCCCTGCTTCAG	1177
QY	231	GGGGTCCGCATGATCCAGTTGAAATCAAGGTAAAGCAGGTGGTACTACTACAAGTAATA	290	Db	483	GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln	502
Db	179	GlyArgThrTyrGlnPro-----TrpGlnPhePhe	188	QY	1178	GGAGAGAGCGCCGAGAAAGAACACTGCTCCAAGTGCCTGCAGAGCAGATCATCCTGGAG	1237
QY	291	ATCCGGAAGAACACTTTTCAGAACTCTTCTGCTCAGGAATCCCTGTTCCAGATGTTCCCAT	350	Db	503	GlyAsnAlaCysArgLysAsp-----Pro	510
Db	189	AlaSerSerLysArgAspCys-----LeuGluArgPhe	199	QY	1238	GAATTGGAGAGGTGCTTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT	1285
QY	351	GGTCCCGAAGAACTA-----GAGGATGCTCCTGCTGTCTCTTTAAGAAA	394	Db	511	ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla	530
Db	200	GlyProGlnThrLeuGluArgIleThrArgAspAspAlaAlaIleCysThrThrGluTyr	219	QY	1286	AAGGNTTCAAAATGGAGTAGAGTATAAAGTGTGAATGC	1324
QY	395	GATTCCCAACCAATGGTGCATATGCCAATTGAAGGGGGCCACACAATGCTATGTATAGAC	454	Db	531	ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys	543

Search completed: April 21, 2003, 12:06:48
Job time : 116 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:51:19 ; Search time 65.5 seconds

(without alignments)
6264.148 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtggaaggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09743237/runat_21042003_112210_1486/app_query.fasta_1.2311
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237@cgn_1_1_111_@runat_21042003_112210_1486 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	11.4	429	T23152	hypothetical prote
2	376	9.8	571	B84585	hypothetical prote
3	358	9.4	603	T08955	hypothetical prote
4	248	6.5	658	F71410	hypothetical prote
5	123	3.2	1700	S08167	Balbani ring 3 pr
6	121	3.2	1737	T00209	MEGF8 protein - hu
7	119.5	3.1	1001	T00532	probable cadmium-t
8	119.5	3.1	1172	F84572	probable cadmium-t
9	119.5	3.1	3010	1 GNVVCJ	genome polyprotein
10	119	3.1	1895	T15881	hypothetical prote
11	118.5	3.1	1801	1 MMRTS	laminin beta-2 cha
12	118.5	3.1	5376	T42215	zonadhesin - mouse
13	118	3.1	2406	2 A54148	odz protein - frui
14	118	3.1	2515	2 S47008	tenascin-like prot

15	117.5	3.1	3507	2 T34513	hypothetical prote
16	117	3.1	497	2 T27827	hypothetical prote
17	117	3.1	3011	1 GNVVC3	genome polyprotein
18	117	3.1	3635	1 T10053	laminin alpha 5 ch
19	116	3.0	3010	1 GNVWTC	genome polyprotein
20	116	3.0	3010	1 A45573	genome polyprotein
21	115.5	3.0	769	1 IJHULM	leukocyte adhesion
22	115.5	3.0	1766	2 A42125	trophozoite cystei
23	115.5	3.0	2195	2 T34264	hypothetical prote
24	114.5	3.0	778	2 T38487	tastin - human
25	114.5	3.0	3759	2 A35085	trithorax protein
26	113.5	3.0	328	2 G89152	protein C24B5.5 [1
27	113	3.0	1820	2 A55494	latent transformin
28	112.5	2.9	574	2 B88465	protein B0244.8 [1
29	112.5	2.9	3133	2 S52093	hemocytin - silkw
30	112	2.9	1599	2 T16210	hypothetical prote
31	112	2.9	3010	1 GNVWTV	genome polyprotein
32	111	2.9	798	2 A40526	integrin beta-7 ch
33	110	2.9	1819	2 A71928	cag island protein
34	109.5	2.9	573	2 A35533	cell surface glyco
35	109.5	2.9	3010	1 S18030	genome polyprotein
36	109	2.9	2180	2 T29764	hypothetical prote
37	109	2.9	2499	1 A30788	mannose 6-phosphat
38	108.5	2.8	1548	2 S34583	serine proteinase
39	108.5	2.8	2325	2 T26063	hypothetical prote
40	108	2.8	1607	2 T02837	long chain fatty a
41	108	2.8	1840	2 T03250	GTI protein - mous
42	108	2.9	5376	2 T42215	zonadhesin - mouse
43	107	2.8	559	2 A57474	extracellular matr
44	106.5	2.8	1786	1 MMHUB1	laminin beta-1 cha
45	106.5	2.8	1797	2 A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T23152
hypothetical protein JC8.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23152
R:Lightning, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19699
A:Accession: T23152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-429 <WIL>
A:Cross-references: EMBL:282274; PIDN:CAB05228.1; GSPDB:GN000022; CESP:JC8.6b
A:Experimental source: Clone JC8
C:Genetics:
A:Gene: CESP:JC8.6b
A:Map position: 4
A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Alignment Scores:
Pred. No.: 4,59e-28 Length: 429
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35
Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 2 Gaps: 8

US-09-743-237-3 (1-2134) x T23152 (1-429)

QY 698 TACTGTGACTGCTTTCAGTGGGACTTTTGCACCACTGCAATTTGTAATAATTGTTGC 757
|||||
Db 182 TyrcysaspCysphealaasnGlyGlupheCysargaspCysasnCysLysaspCysHis 201
|||||
QY 758 AACAACTTCATCATGATTAATTAAGCCATGCTTGTAGTAAT 817
|||||
Db 202 AsnAsnIleGlutyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 221
|||||

QY	287	AATAATCCGGAAGAAGCAACTTTCGCAGAAATCTTTCTTGCTCAGGAATCCCTGTTCGCAAGTTC	346
		: : : : :	
Db	219	AsnGlyValGluLysGlnThrMetGlnHis-----	228
QY	347	CCATGGTCCCAGGAAC TAGAGGATGCCTCGCTGTTCTCTTAAGAAAGATTCACAACCCA	406
		: : : : :	
Db	229	-----AspSerAsnLysGluProGluSerAlaAsnAlaIleProfyrGluValAsnSer	246
QY	407	ATGGTGATATGCCAA-----TTCAAAGGGGGCACACAATGCTATGATAGAC	454
		: : : : :	
Db	247	GlyValIleSerGlnAlaValSerLeuLeuHisArgGlyIleArgArgCysLeuAsp	266
QY	455	-----AATTCTAGAACACAGAGAACTA	475
		: : : : :	
Db	267	PheGluMetProGlyAsnLysGlnThrSerSerGluAsnAsnThrAlaAlaCysGluSer	286
QY	476	AAAGCATCCATTGGTTCTCCTCAGTAT-----	502
		: : : : : :	
Db	287	SerSerArgCysValValProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSerSer	306
QY	503	CAAGATCAAAAATAATTATCTACAGTCAGATGTCCTCAACCAATGACTGCTTTAGTAGGG	562
		: : : : : :	
Db	307	LysAspCysLysThrAsnValThrGlnAspTyrSerCysSerAlaAsnIleGlnValGly	326
QY	563	AGATTTTTGGCCAGCATCAACAAAAATTAATCTCATTAACACAACAACATTTGAGGGAGCCTTA	622
		: : : : :	

Db 467 uLysArgTrpAsnAlaAspThrCysLysCysGluCys----- 479
QY 1348 AAATCTCTAGTTTGAAGAAGGATGTTTAGGGGAACATGAGGCTGGCTCTGCACCAACACC 1407
Db 480 -----GlnPr 481
QY 1408 AGGCTCCCTGCATCCCTGGGGCCAGGAGTTTACTCA----- 1445
Db 481 OglyMetProGlu---GlyCysGlyLys-GlnThrTrpLysSerAspLysCysLysC 500
QY 1446 --GAGCTCTCTGAAGTGGCAACCCATGCCCCCTTTCTGAGGAGTGCATGCCCTGA 1503
Db 500 yslCysSerProThrIleThrCysGlnAlaProGlnIleLeuAspLeuAsnThrCysG 520
QY 1504 GCATTGTTTGT-----CTGGCCACAGAGAGAGCTGGGTTCCCATAGTCT 1551
Db 520 lUcysLysCysProValAsnMetLeuAlaGlnLysGluLysCysLysSerProArgGlnT 540
QY 1552 GG-----GAGAGTGTCTGCAGGGCGGGGAGGAGGAGAGAGGCGCTCGGAGAGCT 1602
Db 540 rPThrAspSerLysCysLeuCysGluCysSerThrThrProAlaThrCysGluGlyLysG 560
QY 1603 CACTCTGGTCTGCTCTCTCAGAGATGTTGCTCTGGAGGC 1646
Db 560 lNThrTrpCysGlyGluAlaCysGlnCysIleCysProGlyGly 574

RESULT 6
T00209
MEGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1737 <NAK>
A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308
A:Experimental source: Brain; clone HGI392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Alignment Scores:
Pred. No.: 0.0907 Length: 1737
Score: 121.00 Matches: 125
Percent Similarity: 29.04% Conservative: 60
Best Local Similarity: 19.62% Mismatches: 187
Query Match: 3.16% Indels: 265
DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x T00209 (1-1737)

QY 11 CAAGCGAAGCTCGCGGGGGGACAGCGCGGGAGCTCTCGGGAGTACCCCGG 70
Db 614 GlnGlyAlaLysArgAspArgMetArgAsnValArgGlySerSerArgGlyLeuGln 633
QY 71 GATCCAGAGCTCAGCGGCTGGAGAGCTCGCGTCTCTGAGGCCGCCGCCGCCGCC 130
Db 634 ValPro-----GlyGluGlnProGlySerTrpGlyPheArgGluValArg 648
QY 131 CTGCAACGTGCACTT---CCTGTCTCTGCTGCTACCCCGGCACCGCGCGGGTGT 187
Db 649 LysLysMetAlaLeuTrpAlaAlaLeuAlaGlyThr-----GlyGlyPhe 663
QY 188 TTGCCCCCTGGCGCTGGGTCTCT---GGAAGGAGGCTCTCCACCGCGCGCGCATGA 244
Db 664 LeuGlu-----GluIleSerProHisLeuLysGluProArgProArgLeuPheHisAla 681
QY 245 TCCCAAGTTGAATCAAGGTAGCAGGTGGTACTACTACA-----AGTAATAT 292

Db 682 SerAlaLeu-----LeuGlyAspThrMetValValLeuLeuGlyGlyArgSerAsp 697
QY 293 CCGGAAGAAGCAACTTTTCAGAAATCTTTCGTTCAGGAATCTGTGTGCAAGTTCCCATGG 352
Db 698 ProAspGluPheSerSerAspValLeuLeuTyrglnValAsnCys----- 712
QY 353 TCCCAGGAACTAGAGGATGCCTCTCTCTCTTAAGAAAGATTCACCAACCAATGGTG 412
Db 713 -----AsnAlaTrpLeu 716
QY 413 ATATGCCAATTTGAAGGGGACACAAAATGCTATGTATAGACAATTTCTAGACAAGAA 472
Db 717 LeuProAspLeu-----ThrArg--- 722
QY 473 CTAAAAGCACTCCATTTGGTTCTCTCAGTATCAAGATCAAAATATTTATCTACATCAGAT 532
Db 723 -----SerAlaSer 725
QY 533 GTCCCTAAACCAATGACTGCTTTTAGTGGGAGATTTTTCGCCAGCA--TCACAAAATTA 589
Db 726 ValGlyProProMetGluSerValAlaHisAlaValAlaValGlySerArgLeu 745
QY 590 AATCTATTACACAACAACTTGGAGGCGCTTACCATCGGTAGTCAACGGGTCTGTCTTC 649
Db 746 -----TyrIleSerGlyGlyPheGlyValAlaLeuGlyArgLeuLeu 760
QY 650 CCTCGGGATCAACTCTTCCAGGACCAACAAAATAAATTTTGGCTGGGTACTGTGACTGC 709
Db 761 -----AlaLeuThrLeuProAspPro-----CysArgLeu 771
QY 710 TTTGCCAGTGGGACTTTTTCACAACAC--TGCAAT--TGTAATAATTTGCTGC 757
Db 772 LeuSerSerProGluAlaCysAsnGlnSerGlyAlaCysThrTrpCysHisGlyAlaCys 791
QY 758 -----AACAACTTGCAT----- 769
Db 792 LeuSerGlyAspGlnAlaHisArgLeuGlyCysGlySerProCysSerProMetPro 811
QY 770 -----CATGATATTGAACGGTTTAAAGCCATTAAGCCATGCTTCTGGTGAATCCA 820
Db 812 ArgSerProGluGluCysArgArgLeuArgThrCysSerGluCysLeuAlaArgHisPro 831
QY 821 GAAGCTTTCCAGCCAAATAATTGGGAAGGCCAATTGGGCAATGTCAAGCCACGACACAAC 880
Db 832 ArgThrLeuGlnPro-----GlyAspGlyGluAlaSerThr-----ProArgCysLys 847
QY 881 AAAGGTGCAACTGCAGGAGTCAAGCTGCCTG-----AAG 916
Db 848 TrpCysThrAsnCysProGluGlyAlaCysIleGlyArgAsnGlySerCysThrSerGlu 867
QY 917 AATTACTCGAGTGTATGAGGCCCAATAATTATGTCTTCTTCTATTGCTG----- 964
Db 868 AsnAspCysArgIleAsnGlnArgGluValPheTrpAlaGlyAsnCysSerGluAlaAla 887
QY 965 -----AAATGCATTGGTTGCAAAAT 985
Db 888 CysGlyAlaAlaAspCysGluGlnCysThrArgGluGlyCysCysMetTrpThrArgGln 907
QY 986 TATGAAGAAGCCAGACGAAAGACACTAATGAGCATG---CCAAACTAC----- 1033
Db 908 PheLysArgThrGlyGluThrArgArgIleLeuSerValGlnProThrTrpArgTrpThr 927
QY 1034 -----ATGCAGACTGGAGGTTTGAAGGCCAGCATTTACCTG 1069
Db 928 CysPheSerHisSerLeuLeuAsnValSerProMetProValGluSerSerProLeu 947
QY 1070 CCA---CCAACCAAAATTTTCAGGACTTCCAAGATTTCAGTCAAGATAGGCGGCTTCTCTCA 1126
Db 948 ProCysProThrProCysHisLeuLeuPro-----Asn 958
QY 1127 TGCATCTCTCGGAGGTGTGTGGAGGCCACATCGCCTGCTGCTCTCAGGAGAGAG 1186
Db 1127 TGCATCTCTCGGAGGTGTGTGGAGGCCACATCGCCTGCTGCTCTCAGGAGAGAG 1186

Db 959 CysThrSer-----CysLeuAspSerLysGlyAlaAsp 969
Qy 1187 GCCGAGAAAGAACACTGCTCAAGTGCCTGGCAGAGCAGATGATCCTGGAGAAATTTGGA 1246
Db 970 GlyGlyTrpGlnHisCys-----ValTrpSer 978
Qy 1247 AGTGCTTATACAGATTCTCCACACTGAGTTTAAATCTAAGGATTGAAAATGGAGTAG 1306
Db 979 SerSerLeuGlnCysLeuSerProSerTyr----- 989
Qy 1307 AGTATAAGTGTGAATGCATGCTGATTTTGTCTTAGTCTAGATAATCTAGTTTAGAAG 1366
Db 990 ---LeuProLeuArgCysMet----- 995
Qy 1367 GATGTTTAGGGGAACATGAGGCTGGCTGCGAGCAACAACAGGCTCCCTGTCATCCTG 1426
Db 996 -----AlaGly 997
Qy 1427 GGCCAGGAGGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCATGCCCTTTTCTGA 1486
Db 998 GlyCysGlyArgLeuLeuArgGly----- 1005
Qy 1487 GGAGTGATGGCTGAGCATTTGCTGCGCCAGAGGAGAGAGCTTGGGTTCCCATTA 1546
Db 1006 -----ProGluSerCysSerLeuGlyCysAlaGlnAlaThrGlnCysAlaLeu 1021
Qy 1547 GTCCTG-----GGAGAGTGTCTGCGAGGCGGC----- 1573
Db 1022 CysLeuArgArgProHisCysGlyTrpCysAlaTrpGlyGlnAspGlyGlyArg 1041
Qy 1574 -----GGAGGCGAGCAGGCCCTGCGGAG-----AGCTC 1603
Db 1042 CysMetGluGlyGlyLeuSer-GlyProArgAspGlyLeuThrCysGlyArgProGlyAl 1061
Qy 1604 ACTCTGGTGGACTCT-----CCTCTCAGAGATGTTGCTCTGGA 1643
Db 1061 aSerTrpAlaPheLeuSerCysProGluAspGluCysAlaAsnGly 1077

RESULT 7

T00532
Probable cadmium-transporting ATPase (EC 3.6.1.-) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C:Accession: T00532
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A:Reference number: Z14167

A:Accession: T00532
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1001 <R0U>
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176713
A:Experimental source: cultivar Columbia

C:Genetics:
A:Map position: 2
A:Introns: 27/3; 113/1; 159/1; 268/2; 336/1; 403/3; 980/1
A:Note: T20K24.13
C:Superfamily: ATPase nucleotide-binding domain homology
C:Keywords: hydrolase
F:360-504/Domain: ATPase nucleotide-binding domain homology <ATN>

Alignment Scores:
Pred. No.: 0.119 Length: 1001
Score: 119.50 Matches: 110
Percent Similarity: 29.72% Conservative: 57
Best Local Similarity: 19.57% Mismatches: 184
Query Match: 3.12% Indels: 211
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x T00532 (1-1001)

Qy 235 GTCCGCATGATCCCGAGTTGAATCAAGGTAGCAGGTGGTACTACTACAAGTAATATATCC 294

Db 476 IleArgGlnProGlnAlaValLysLeuAlaArg-----Arg 488
Qy 295 GGAAGAAGCAACTTTGCGAGAATCTTCTGCTCAGGAATCTGTTGCAAGTTCCCATGTC 354
Db 489 AlaArgGlyValValGluAsnValCysLeuSerIleIleLeu-Lys----- 504
Qy 355 CCAGGAACATAGAGGATGCCTGCTGCTCTTCTTAAAGAAAGATTCCCAACCAATGGTG-- 412
Db 505 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 519
Qy 413 -----ATATGCCAATTGAAGGGGCACAAATCTATGTATAGACAATCTAGA----- 463
Db 519 pAlaAlaValLeuValAspValGlyThrCysLeuValIlePheAsnSerMetLeuLe 539
Qy 464 -ACAAGAGAAGCACTAAACACTCCCATTTGGTTCCTCAGTATCAAGTCAAAATATTATCT 522
Db 539 uLeuArgGlyLysLysIleGlyAsnLysLysCysTyrArg----- 553
Qy 523 ACAGTCAGATGTCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTGGCAGCATCAAC 582
Db 554 -----AlaSerThrSe 557
Qy 583 AAAATTAATCTCATTTACACAACACTGAGGGAGCCTTACCATCGGTAGTCAAC----- 637
Db 557 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuGl 575
Qy 638 ---GGGTGCTGCTTCCCTCGGATCACTCTCCAGGACCACCAAAATAAATTTGGC 693
Db 575 uAlaGlyLeuLeuThrLysSerGlyAsn----- 584
Qy 694 TGGTACTGTGACTGCTTGGCAGTGGGACTTTTGCACAACACTGCAAT----- 742
Db 585 -GlyGlnCysLysSerSerCysCysGlyAspLysLysAsnGlnGluAsnValValMetMe 604
Qy 743 -----TGT----- 745
Db 604 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLy 624
Qy 746 -----AATAATTGTTGC----- 757
Db 624 sGluGluLysValLysProLeuValLysAspGlyCysSerGlyLysThrArgLysSe 644
Qy 758 -----AACAACTGTCATCATATAT 777
Db 644 rGluGlyAspMetValSerLeuSerSerCysLysSerSerHisValLysHisAspLe 664
Qy 778 TGAACGGTTTAAAGCCATTAGGCATGCTTGGTAGAAATCCAGAAGCTTTCAGCCAAA 837
Db 664 u---LysMetLysGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 683
Qy 838 AATTGGGAAGGCCAATTGGCAATGTCAAGCCCCCAGCACACAAGGGTGCACACTGCAG 897
Db 683 lValAlaLysSerCysCysGluLysProLysGlnGlnValGluSerValGlyAspCysLy 703
Qy 898 GAGTTCAGGCTGCCTGAAG----- 916
Db 703 sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValProValGlnIleIl 723
Qy 917 -----AATTACTGGAGTCTATAGGCCCAAAATATGTTCTTCTATTTCGAATGCAT 972
Db 723 eGlyHisAlaLeuThrHisValGluIleGluLeuGlnThrLysGluThrCysTyrThrSe 743
Qy 973 TGGTTGCAAAAATTTATGAAGAAAGCCAGAAAGACACTAATAGCATGCCAAAC-- 1030
Db 743 rCysCysAspSerLysGlyLysValLysGluThrGlyLeuLeuSerSerGluAsnTh 763
Qy 1031 -----TACATGCAGACTGGAGTTTG-----GAAGCAGCCCA 1062
Db 763 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl 783
Qy 1063 TTACCTCCCAACCAAGAAATTTTCAGGACTTCCAAGATTCAAGTTCAGTCACGATAGGCGCCTTC 1122

Db 783 uAnMetGlyThrVallysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 803
QY 1123 CTCATCATCTCTCTGGAGGTGGTGGAGGCCACATCGCGTCTGCTGCTCAGGAGA 1182
Db 803 rGlyGluLeuThr-----LeuAlaSerGluG1 812
QY 1183 AGAGCGCGAGAAAGAACTGCTCCAG-----TGCCTGGCAGAGCAGATGATCTCGA 1236
Db 812 uGluThrAspGlnAspCysSerGlyCysValAsnGluGlyThrValLysG1 832
QY 1237 GGAATTGGAGGTCTTATCATCAGATTCTCCACATGATTTAAATCTAAGGATTGAA 1296
Db 832 nSerPheAspGluLysHisSerValLeu-----ValGluLysGluGlyLeuAs 849
QY 1297 AATGGAGTAGAGTATAAGTGTCAATGCTGTTGATTTGCTCTAGCTAGAAATCTCTA 1356
Db 849 pMetGluThrGlyPheCysAspAlaLysLeuValCys----- 862
QY 1357 GTTTAGAAAGGATTTAGGGGAACATGAGGCTGCTGTCAGCAACAACAGGCTCCCC 1416
Db 862 ----- 862
QY 1417 TGCATCCCTGGGCCAGGGAGTTTACTCAGAGCTCTGAGATGTGGCAACCCATGCC 1476
Db 863 ----- 866
QY 1477 CCTTTCTGAGAGGTGCATGCGCTGAGCATTTGTCTGCGCCAGAGGAGAGAGC 1533
Db 867 -----GluGlyGluValLysGlnCysArgLeuGluLeuLysLysGluGlu-- 882
QY 1534 TTGGGTCCCATAGTCTCTGGAGAGTGTCTGACGGCGCGGAGGAGCAGAGGCCCTG 1593
Db 883 -----HisCysLysSerGlyCysGlyGluGluLeuGlnTh 895
QY 1594 CGGAGAGTCACTCTGCTGCTCTCTCTCTCAGCAAGT-----TGCCTGAGGCT 1647
Db 895 rGlyGluLeuThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 915
QY 1648 GC 1649
Db 915 ys 915

RESULT 8
F84572

probable cadmium-transporting-ATPase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84572

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: F84572

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1172 <STO>

A:Cross-references: GB:AE002093; NID:94210504; PIDN:ADI2041.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g19110

A:Map position: 2

Alignment Scores:

Pred. No.:	0.12	Length:	1172
Score:	119.50	Matches:	110
Percent Similarity:	29.72%	Conservative:	57
Best Local Similarity:	19.57%	Mismatches:	184
Query Match:	3.12%	Indels:	211
DB:	2	Gaps:	23

US-09-743-237-3 (1-2134) x F84572 (1-1172)

QY 235 GTCCGCATGATCCCACTTGAATCAAGCAAGGTGGTGGTACTACAGTAATAATCC 294
Db 639 IleArgIleProGlnAlaValLysLeuAlaArg-----Arg 651
QY 295 GGAAGAAGCAACTTTTCAGAAATCTTCTGCTCAGGAATCTCTGTTGCAAGTTCCCATGGTC 354
Db 652 AlaArgArgLysValValGluAsnValCysLeuSerIleIleLeu-Lys----- 667
QY 355 CCAGGAACATAGAGGATCGCTCTCTTCTTAAGAAAGATCCCAACCAATGGTG-- 412
Db 668 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProIleIleTr 682
QY 413 ----ATATGCCAATTGAAAGGGGCACACAAATGCTATGATAGAAATCTAGA----- 463
Db 682 pAlaAlaValLeuValAspValGlyThrCysLeuLeuValIlePheAsnSerMetLeuLe 702
QY 464 -ACAAGAGAACTAAAGCACTCCATTTGGTCTCTCAGTATCAAGATCAAAATATATCT 522
Db 702 uLeuArgGluLysLysIleGlyAsnLysLysCysTyArg----- 716
QY 523 ACAGTCAGATGTCCTTAAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAAC 582
Db 717 -----AlaSerThrSe 720
QY 583 AAAATAAATCTCATTACACAACTTGAGGAGCCTTACCATCGGTAGTCAAC----- 637
Db 720 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyValValAspLeuG1 738
QY 638 ----GGTCTGCTTCCCTCGGGATCACTCTTCCAGGACCACCAAAATACTTTGGC 693
Db 738 uAlaGlyLeuLeuThrLysSerGlyAsn----- 747
QY 694 TGGTACTGAGCTGCTTTCCAGTGGGACTTTTCCAACTGCAAT----- 742
Db 748 -GlyGlnCysLysSerSerCysGlyAspLysLysAsnGlnGluAsnValValMetMe 767
QY 743 -----TGT----- 745
Db 767 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLys 787
QY 746 -----AATAATTGTTGC----- 757
Db 787 sGluGluLysValLysProLeuValLysAspGlyCysSerSerGluLysThrArgLysSe 807
QY 758 -----AACAACCTGCATCATGATAT 777
Db 807 rGluGlyAspMetValSerLeuSerSerCysLysLysSerSerHisValLysHisAspLe 827
QY 778 TGAACGGTTTAAAGCCATTAAAGCATGCTCTTGGTAGAAATCCAGAGCTTTCCAGCCAAA 837
Db 827 u----LysMetLysGlyGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 846
QY 838 AATTGGGAAGGCCAATTGGCAANTGTCAAGCCCGCAGCACAAAGGGTGCAACTGCAG 897
Db 846 lValAlaLysSerCysCysGluLysProLysGlnValGluSerValGlyAspCysLys 866
QY 898 GAGGTCAGGCTGCTCAAG----- 916
Db 866 sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValProValGlnIleIl 886
QY 917 ----AATTACTGCGAGTGTATGAGGCCCAAAATATGTTCTTCTTATTTGCAATATGCAT 972
Db 886 eGlyHisAlaLeuThrHisValGluLeuGluGlnThrLysGluThrCysLysThrSe 906
QY 973 TGGTTCAAAAATATGAAGAAAGCCAGCAACAAAGACACTAATGAGCATGCCAAC-- 1030
Db 906 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuSerSerGluAsnTh 926
QY 1031 ----TACATGCCAGCTGGAGTTTG-----GAAGCCAGCCA 1062
Db 926 rProTyLeuGluLysGlyValLeuLysAspGluGlyAsnCysLysSerGlySerG1 946
QY 1063 TTACCTGCCACCAACGAAATTTTTCAGGACTTCCAGATTCAGTCACGATAGGGCGCTTC 1122

Db	946	uAsnMeGlyThrValylsGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh	966
Qy	1123	CTCATGCATCTCCTGGAGAGTGGTGGAGCCACATGCGCTGCTGTGCTTCAGGAGA	1182
Db	966	rGlyGluIleThr-----LeuAlaSerGluGI	975
Qy	1183	AGAGCGCGAGAAAGACACTGCTCCAG-----TGCTGGCAGAGCAGATGATCTCGGA	1236
Db	975	uGluThrAspAspGlnAspCysSerSerGlyCysValAsnGluGlyThrValLysGI	995
Qy	1237	GGAATTGGAAAGCTGCTTATCACAGATTCTCCACACTCAGTTTAAATCTAAGGATGAA	1296
Db	995	nSerPheAspGluLysLysHisSerValLeu-----ValGluLysGluGlyLeuAs	1012
Qy	1297	AATGGAGTAGAGTATAAAGTGAATGCATGTTGATTTGCTTAGTCTAGAAATCTCTA	1356
Db	1012	pMetGluThrGlyPheCysCysAspAlaLysLeuValCys-----	1025
Qy	1357	GTTTAGAAGGATGTTTAGGGGAACATGAGGCTGCTCGCAGCAACACCAGCTCCCC	1416
Db	1025	-----	1025
Qy	1417	TGCATCCCTGGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCATGCC	1476
Db	1026	-----CysGlyAsnThr-----	1029
Qy	1477	CTTTTCTGAGAGNGCATGGCCTGAGCATGT---TTGTCTGGCCCGAGGAGAGAGC	1533
Db	1030	-----GluGlyGluValylsGluIleGlnCysArgLeuGluIleLysLysGluGlu--	1045
Qy	1534	TTGGTTCCTCATAGTCTCTGGGAGAGTGCTCGAGGCGCGGAGGCGAGCAGAGCCCTG	1593
Db	1046	-----HisCysLysSerGlyCysCysGlyGluGluIleGlnTh	1058
Qy	1594	CGGAGACTCACTCTGGTCTCTCTCAGAGATGT-----TGCTCTGGAGGCT	1647
Db	1058	rGlyGluIleThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC	1078
Qy	1648	GC	1649
Db	1078	vs	1078

RESULT 9
GNMYCJ

genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hljikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K. U.S.A. 87, 9524-9528, 1990
Proc. Natl. Acad. Sci.
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; PMID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:CROSS-references: GB:D90708; NID:G221610; PIDN:BAAL4233.1; PID:G221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variability
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KA2>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine protease
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <MEE>

```

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepacivirin #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (p-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196_209_234_250_305_325_417_423_430_448_532_556_576_623_645_1213_1255_2041_2077_2224

```

US-09-743-237-3 (1-2134) x GNWVCJ (1-3010)

Qy	65	CCCCGGGATCCCAGAGCTCAGCGCGCTGGAGGAGCTGCGGCTCTCGCAGGC-----	115
Db	1217	ProAlaValProGlnThrPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLys	1236
Qy	116	-----CCCCACGCGCCGCCCTGCAACGTGCACCTTCCTGTCTCCTCGCTCCTACC	163
Db	1237	SerThrLysValProAlaAlaTyrAlaAlaGln-----GlyTyrLysValLeuValLeuAsn	1255
Qy	164	CGCGCACGACGCGCGGGGTGTTTGGCCCTGGGCGCGCGCT-----GGTCTCGCG	214
Db	1256	-----ProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGly	1273
Qy	215	AAGAGCCTCCCAACCGCGGCGCCGATGCCAGTTCAGTTCGAATCAAGTAAGCAGGTGGT	274
Db	1274	IleGluProAsnIleargThrGlyVal-----Arg	1283
Qy	275	ACTACTACAGTAATAATCCGGAAGAACCACTTTGCCAGAACTCTCTTGGCTCAGAAATCC	334
Db	1284	ThrIleThrThrGlyGlyPro-----IleThrTyrSerThr	1295
Qy	335	TGTTGCAAGTTCCTCATGGTCCCGAGGAACCTAGAGGATGCTCCTGCTGTCTCTTAAGAAA	394
Db	1296	TyrCysLysPhe-----LeuAlaAspGlyGlyCysSerGlyGlyAlaTyr	1310
Qy	395	GATTCCACCCAAATGGTGATATGCCAATTGAAGGGGGCAGCAAACTCTATGTATA---	451
Db	1311	Asp-----IleIleIleCysAspGluCysHisSerThrAspSerThrIleLeu	1327
Qy	452	-----CACAAATCTAGAACACAGAGAACTAAAGCACTCCATTCGTT	493
Db	1328	GlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAla	1347
Qy	494	-----CCTCAGTATCAAGATCAAAATAAT	517
Db	1348	ThrAlaThrProGlySerIleThrValProHisProAsnIleGluGluValAlaLeu	1367
Qy	518	TATCTACAGTCAGATGTCCT-----AAACAATGACTGCTTTAGTAGGG	562
Db	1368	SerAsnThrGlyLeuIleProPheTyrGlyLysAlaIleProIleGluAlaIleLysGly	1387
Qy	563	AGA-----TTTGTGGCCAGCATCAACAAAATAATCTCATTCACACACACTT	610
Db	1388	GlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeu	1407
Qy	611	GAGGGA-----GCCTTACCATCGGTA-----GTCAACGGGTGCTGCTTCCCC	652
Db	1408	ThrGlyLeuGlyLeuAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIlePro	1427
Qy	653	-----TCGGGATCAACTCTCCAGGACACCAAAAAATACTTTGGCTGGGTACTGTGACTGC	709
Db	1428	ThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyPhe	1444

```
QY 710 TTGCGCAGTGGGACTTTTCCAAACAACTGCAATTTGTAATAATTTGTCACAACTTG--- 766
Db 1445 -----ThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrVal 1462
QY 767 -----CATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGTGCTT 808
Db 1463 AspPheSerLeuAspProThrPheThrIleGluThrThrThrLeuProGlnAspAlaVal 1482
QY 809 GGTAGAATCCAGAAGCTTTCAGCCCAAAATTTGGGAAGGCCAAATTTGGGC----- 859
Db 1483 SerArgAlaGlnArg-----ArgGlyArgThrGlyArgGlyArgSerGlyIleTyrArg 1500
QY 860 AATGTCAGCCCGCCAGCACAACAAAGGTGCAACTGCAGGAGGTCCAGCTCCCTCAAGAAAT 919
Db 1501 PheValThrProGlyGlu-----ArgProSerGlyMetPheAspSer 1514
QY 920 -----TACTGCGAGTCTATGAGGCCCAATTTATGTGTTCTTCTATTTCCAATGCATT 973
Db 1515 SerValLeuCysGluCysTyrAspAla----- 1523
QY 974 GGTGCAAAATTTATGAAGAAACCCAGAACGAAAGACACTAATGAGCATGCCAAACTAC 1033
Db 1524 GlyCysAlaTrpThrGluLeuThrPro---AlaGluThrSerValArgLeuArgAlaTyr 1542
QY 1034 ATGCACTGGAGTTTGGAA----- 1054
Db 1543 LeuAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPhe 1562
QY 1055 -----GCCACCATTTACTGCCACCAACAGAAATTTTCAGGA----- 1090
Db 1563 ThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsn 1582
QY 1091 TTCCCAAGATTCACTC-----ACGATAGGGGGCTTCCCTCATCATCTCCTGG 1138
Db 1583 LeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProPr 1602
QY 1139 GAGGTGG-----TGGAGGCCACATGCGCTGCTGCTCAGGAGGAAGAGGCC 1189
Db 1602 oSerTrpAspGlnMetTrpLys-----CysLeuIleArgLeuLysProTh 1617
QY 1190 GAGAAAGAACACTGCTCCCAAGTCCCTGGCAGACAGCATGATCTGGAGGAATTTGGAAGG 1249
Db 1617 rLeuHisGlyProThrProLeuLeuTyrArgLeu-----G 1629
QY 1250 TGCTTATCACAGATTCTCCACACTGACT----- 1277
Db 1629 yAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIleMetAlaCysMe 1649
QY 1278 -----TTAAATCTAAGGATTGAAATGGAGTAGAGTATAAGTGTGAATGCATG 1327
Db 1649 tSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyValLeuAla 1669
QY 1328 TTGATTTTGTCTAGTCTAGAAATCTCTAGTTTGAAGAGATGTTAGGGGAACATGAGG 1387
Db 1669 aLeuAlaAlaTyrCysLeuThrThrGlySerValIleVal-----G 1684
QY 1388 CTGGCTCTGCAGCAACACAGCTCCCTGTCATCCCTGGGCCAGGGAGCTTTACTCAGA 1447
Db 1684 yArgIleLeuSerGlyArgProAlaValIleProAspArgGluValLeuThrGlnG 1704
QY 1448 GCTCTCT-----GAAGATGTGGCAACCCATGCCCCCTTTCTGAGGAGGTGTCATGCCCT 1501
Db 1704 uPheAspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGln-----GlyMe 1722
QY 1502 GACCATTTGTTCTGCGCCAGAGGAGAGAGCTTTGGTTCCTCCCATAGTCTCTGGGAGAGTGT 1561
Db 1722 tGlnLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-----LeuGlnThrAl 1739
QY 1562 CTCAGGGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1619
Db 1739 aThrLysGlnAlaGluAla-----AlaAlaProValValGluSerLysTrpArgAlaLeu 1757
```

RESULT 10

T15881

hypothetical protein D1044.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15881

R:Pauley, A.

submitted to the EMBL Data Library, June 1994

A:Description: The sequence of C. elegans cosmid D1044.

A:Reference number: Z18423

A:Accession: T15881

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1895 <PA>

A:Cross-references: EMBL:U00065; NID:g495681; PID:g495684; PIDN:AAA50735.1; CESP:D104

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:D1044.3

A:Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2.

Alignment Scores:

Pred. No.: 0.133 Length: 1895

Score: 119.00 Matches: 107

Percent Similarity: 32.69% Conservative: 80

Best Local Similarity: 18.71% Mismatches: 178

Query Match: 3.11% Indels: 207

DB: 2 Gaps: 30

US-09-743-237-3 (1-2134) x T15881 (1-1895)

QY 31 GCAGACGACGACGGGGAGCTCTCGGGGAGTACCCGGGATCCCGGATCCAGAGCTCAGCGCGC 90

Db 732 SerSerLeuAsnSerSerThrSerSerTyrLysArgProGlyArgSerSer-LysThrSe 751

QY 91 TGA---GGACGTGGCTCTCTCAGCGCCCGCAGCGCCCGCTGCAAGCTGACTTCC 147

Db 751 rGlySerGlyLysLeuProIleAsnProAlaProArgArgAsnHisSer----- 768

QY 148 TGCTCTGCTGCTACCCGCGCACCGCGCGGGTGTTCCTCCCTGGCGGCGCTGGG 207

Db 769 ---MetSerAlaGlyGluThrProArgProSerSerIleValCysLeuProArgIleTh 787

QY 208 T-----CCTCGAAG----- 217

Db 787 rAspLysTyrSerValMetPheProSerAlaProSerAlaIleProSerArgArgVa 807

QY 218 -----GAGCTCCACCGCGGCTCGCATGATCCAGTTGAAA---TC 258

Db 807 lGlnThrCysSerThrGluHisProAlaArgSerSerSerSerThrGluLeuLysValSe 827

QY 259 AAGSTAAGCAGGTGTACTACTACAGTAATAAT----- 292

Db 827 rLysGlnSerAspGlyLeuThrValSerSerAsnValLeuArgGlnIleGlnAspProVa 847

QY 293 -----CCGGAAGAAGCAACTTTG-- 310

Db 847 lLeuThrIleLeuAsnGlnIleHisArgIleLeuValValThrAspLysGluThrIleSe 867

QY 311 -----CAGATCTCTTCTGCTCA 327

Db 867 rThrSerMetAsnHisGlnArgArgLeuValGluMetPheArgLysAsnLeuLeuGlyAr 887

QY 328 GGAATCC-----TGTTGCAAGTTCCCA-- 349

Db 887 gGluAsnAspAlaValGlnMetLysThrHisLeuArgLysLeuAlaGlnSerProGl 907

QY 350 -----TGGTCCCGAAGACTAGAGGATCCCTCTG 378

Db 907 uGluIleGlnMetAsnLeuGlyPheSerAspPheArgProValLeuValGlnSerHisII 927

QY 379 CTGTTCTCTTAAAGAAAGATTCCAAACCAATGGTGATATGC-----CAATTGAAAGGGG 432

Db 927 eAsnGlyTyrGlnLysAspGlnLysValThrLysIleThrTyrGluGlnLeuSerAlaCy 947

QY 433 CACACAAATGCTATGATAGACAAT-----TCTAGAACAAAGAGA 471
Db 947 stleGlucysLeuIleAlaGluAsnProAlaAlaLysHisValProHisArgThrArgAl 967
QY 472 ACTA-----AAAGCACTCCATTGTTCTCAGTATCAAGATCAAAATAATA 519
Db 967 aValValIleLeuArgAspLeuValLeuGlyGlnTy----- 981
QY 520 TCTACAGTCAGATGCCCTAAACCAATGACTGCTTTAGTAGGAGATTTTCCAGCATC 579
Db 982 -----ValAsnMetLeuValProThrIl 989
QY 580 AACAAAAATAATCTCATACACAACTTCGAGGAGCCTTACCATCGGTAGTCAACGG 639
Db 989 eThrThrTy-----ValValIleArgGlnIleHisValSerLeuAlaAlaIleLeuValTy 1008
QY 640 GTCTGCTTCCCTCGGATCAACTCTCCAGGACCAACAAATAAATACTTTGGTGGGTA 699
Db 1008 rThrGluTyGluCysGlySerAsn-----SerSerProGlnValSerAlaSerGlyG1 1026
QY 700 C-----TGAGCTGCTTTCCAGTGGGACTTTTCAACAACTG 738
Db 1026 nValValThrCysSerThrAsnThrGlnCys-----AlaSerGlyTyThrCysAsnAsn-- 1044
QY 739 CAATTGTAATAATGTTGCAACAACCTGCATCATGATATTAACCGTTTAAAGCCATTAA 798
Db 1045 -----GlyAlaCysCysProAsnThrAsnSerAsnThrCysSerSerAsnGlyAsnAs 1062
QY 799 GCGATGCTTGTGAATAATCCAGAGCTTCCAGCCAAAATTTGGGAAGGCCAATTGGG 858
Db 1062 nGlyCysLeu-----AlaGlyGlnThrMetVal-----AsnGlyGlnCysTy 1076
QY 859 CAATGCTCAAGCCCGCAGCACAAAGGGTGCACCTGAGGAGG----- 901
Db 1076 rAsn-----SerValAsnIleGlySerAlaCysGlnSerThrGlnGlnCysLeuG1 1093
QY 902 -----TCAGCTGCTGAGAAATTAATCGAGTGCTATGAGGCCCAAAATATGTGTCCTC 957
Db 1093 yGlySerGlnCysGlnAsnAsnIleCysGlnCysTySerGlyTyValAsnValAsnG1 1113
QY 958 TATTTCG----- 966
Db 1113 nGlnCysValIleSerAsnGlyLeuAsnCysGlnLeuGlyThrValSerTyAsnSerG1 1133
QY 967 ATGCATT-----GGTTGCAAAAATATGAAGAAAGCCAGCAACG 1005
Db 1133 nCysIleThrLeuAlaSerProGlyGlnAsnCysGlnThrSerSerGlnCysIleAspAs 1153
QY 1006 AAAGACACTAATGACATG-----CCAACTACATGCAGACTGGAGGTTT 1050
Db 1153 nSerValCysMetAsnGlnMetCysThrCysAsnAsnAsnTyArgLeuValTyGly-- 1172
QY 1051 GGAAGGCCCATTAACCTGCCACCAACCAATTTTCAGGACTTCCCAAGATTCAGTCACGA 1110
Db 1173 -----TyrCysValProIleThr-----SerSerIleCysGlnGlnThrGlnTh 1187
QY 1111 TAGGGGGCTTCTCATCATGCTCTCGGAGGTGGTGGAGCCACATGC----- 1159
Db 1187 rLeuValAsnAsnGlnCysValLeuLeuSerIleValGlyGlnThrCysIleAlaAsnG1 1207
QY 1160 -----CCCTGCTGCTTGTCTCAGGGGAGAGA 1185
Db 1207 nGlnCysValGlyIleAlaMetCysAsnSerGlyThrCysGlnCysThrAsnGlyAlaTh 1227
QY 1186 GCGCCGAGAAGAACTGC-----TCCAAGTGCCTGGCAGAGCAGCATAT 1230
Db 1227 rAlaMetTyGlyTyCysIleSerSerSerSerSerCysAsnSerAsnGlnValSe 1247
QY 1231 CCTGGAGGAATTTGGAAGGTGCTTATCACAGATT 1264
Db 1247 rIleAsn-----GlyMetCysTyAsnThrVal 1256
RESULT 11

MMRTS
laminin beta-2 chain precursor - rat
N:Alternate names: laminin chain B3; S-laminin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C:Accession: S03539
R:Hunter, D.D.; Shah, V.; Merlie, J.P.; Sanes, J.R.
Nature 338, 229-234, 1989
A:Title: A laminin-like adhesive protein concentrated in the synaptic cleft of the ne
A:Reference number: S03539; MUID:89159410; PMID:2922051
A:Accession: S03539
A:Molecule type: mRNA
A:Residues: 1-1801<HUN>
A:Cross-references: EMBL:X16563; NID:g57250; PIDN:CAA34561.1; PID:g57251
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:
A:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-1801/Product: laminin beta-2 chain #status predicted <MAT>
F:36-285/Domain: VI <DOM6>
F:286-555/Domain: V <DOM5>
F:286-347/Domain: laminin-type EGF-like homology <LE01>
F:350-410/Domain: laminin-type EGF-like homology <LE02>
F:413-470/Domain: laminin-type EGF-like homology <LE03>
F:473-522/Domain: laminin-type EGF-like homology <LE04>
F:525-555/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:556-784/Domain: IV <DOM4>
F:786-831/Domain: laminin-type EGF-like homology <LE06>
F:788-1196/Domain: III <DOM3>
F:834-877/Domain: laminin-type EGF-like homology <LE07>
F:880-927/Domain: laminin-type EGF-like homology <LE08>
F:930-986/Domain: laminin-type EGF-like homology <LE09>
F:989-1038/Domain: laminin-type EGF-like homology <LE10>
F:1041-1095/Domain: laminin-type EGF-like homology <LE11>
F:1098-1143/Domain: laminin-type EGF-like homology <LE12>
F:1146-1190/Domain: laminin-type EGF-like homology <LE13>
F:1197-1412/Domain: II <DOM2>
F:1197-1412/Region: heptad repeats
F:1413-1445/Domain: alpha <ALP>
F:1446-1801/Region: heptad repeats
F:1446-1801/Domain: I <DOM1>
F:45-50/Disulfide bonds: #status predicted
F:251,371,1088,1252,1311,1351,1502/Binding site: carbohydrate (Asn) (covalent) #statu
F:1193,1196,1800/Disulfide bonds: interchain #status predicted
Alignment Scores:
Pred. NO.: 0.147 Length: 1801
Score: 118.50 Matches: 110
Percent Similarity: 27.02% Conservative: 54
Best Local Similarity: 18.12% Mismatches: 193
Query Match: 3.10% Indels: 250
DB: 1 Gaps: 29
US-09-743-237-3 (1-2134) x MMRTS (1-1801)
QY 76 CAGAGCTCAGCGCTGGAGGACGTGCGGCTCTGCGAGGCCCGCGCGCTGCA 135
Db 633 GlnValProGluGlnTrpAlaGluLeuValGlnArgProGlyProValSer 652
QY 136 ACGTGCATCTCTGCTCGCTGCTACCCGCGCACCCGCGCGGGGTGTTTGGCCCT 195
Db 653 AlaHisSerProCysGlyHisValLeuProArg----- 663
QY 196 GGGGGCGCTGGTCTCGGAGGAGGAGCCCTCCACCGCGGCTCCGATCATC----- 246
Db 664 -----AspAspArgIleGlnGlyMetLeuHisProAsnThrArgValLeuValPhePro 681
QY 247 ---CCAGTTGAATCAAGGTAAGCAGGTGGTACTACTACAAG-----TAATAA 291
Db 682 ArgProValCysLeuGluProGlyLeuSerTyrlsLeuLysLeuLysLeuThrGlyThr 701

```
QY 292 TCCGGAAGAGCAAC---TTTCAGAAATCTTCTGCTCAGGAATCCTGTTG---CAAGTT 345
|||||:|||||
Db 702 GlyArgAlaHisProGluThrProTyrSerGlyLeuLeuLeuLeuLeuLeuLeu 721
QY 346 CCATGTCCTCCAGGAAGACTAGAGATCCCTCGCTGCTGCTCTCTTAAGAAAGATTCACACCC 405
|||:|||||
Db 722 ValLeuGlnProHisVal-----LeuMetLeu-----GluMetPheSerGly 735
|||||
QY 406 AATGGTGATATGCCAATTAAGAGGGGCACACAAATGCTATGATAGACAATCTTAGAAC 465
|||||
Db 736 GlyAspAlaAlaLeuGluArgArgThrPheGluArgTyrArg----- 751
QY 466 AAGAGAACTAAAGACACTCCATTTGCTTCTCAGTATCAAGATCAAAATAATATATAC 525
|||:|||||
Db 751 ----- 751
QY 526 GTCAGATGTCCTTAACCAATGACTGCTTTAGTAGGGAGATTTTCCAGCATCAACAA 585
|||||:|||||
Db 752 -----CysHisGluGluGlyLeu-MetProSerLysThrPr 763
QY 586 ATTAATCTCATTAACACAACTTGAGGAGCGCTTACCATCGGTAGTCAACGGGTCGCG 645
|||||:|||||
Db 763 OleuSer-----GluAlaCysValProLeuLeuLeuLeuLeuLeuLeuLeu 777
|||||
QY 646 TTTCCCTCGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGA 705
|||||
Db 777 rLeuValTyrAsnGlyAlaLeuPro-----CysGln 787
QY 706 CTGCTTTGCGAGTGGGAGCTTTTGCACAACTGCAATTTGTAATTAATTTGTTGCAACA 765
|||||:|||||
Db 787 nCysAspProGlnGlySerLeuSerSerGluCysAsnProHisGly----- 802
QY 766 GCATCATGATATTGAACGCTTTAAAGCCATTAAGGCATGCTTGTGTAGAAATCCAGAAGC 825
|||||:|||||
Db 803 -----GlyGlnCysArgCysLysProGlyValValGlyArgArgCysAspAl 818
QY 826 TTTCCAGCAAAATTTGGAAGGGCCAAATTTGGCAATGTCAAGCCCCCAGCACAAACAAG 885
|||||:|||||
Db 818 a-----CysAlaThrGlyTyrTyrGly---PheGlyProAlaGlyCysGlnAl 833
QY 886 GTGCAACTGCAGAGTGCAGGTGCTGCTGAAGAAATTAATCTCGAGTGTGATGAGGCCCAAT 945
|||||:|||||
Db 833 aCysGlnCysSerProAspGlyAlaLeuSerAlaLeuCysGluGlyThrSerGlyGlnC 853
|||||:|||||
QY 946 TATGCTCTTCTCT----- 958
|||||
Db 853 sLeuCysArgThrGlyAlaPheGlyLeuArgCysAspHisCysGlnArgGlyGlnTrpG 873
QY 959 -----ATTGCAAA----- 967
|||||
Db 873 yPheProAsnCysArgProCysValCysAsnGlyArgAlaAspGluCysAspAlaHisTh 893
QY 968 -----TGCATTTGTTGCAAAATTAATCAAGAACCCAGAACCAACTAATGAG 1020
|||||:|||||
Db 893 rGlyAlaCysLeuGlyCysArgAspThr----- 902
QY 1021 CATGCCAACTACATGCAGACTGGAGGT-----TTGGAAGGAGCA 1062
|||||:|||||
Db 903 -----ThrGlyGlyGluHisCysGluArgCysIleAlaGlyPheHi 916
QY 1063 TTACCTGCCACCAAGAAATTTTCAGGACTTCCAAGATTCAGTACGATAGCGGCTTC 1122
|||||:|||||
Db 916 sGlyAspProArgLeuProTyrGlyGlyGlnCysArg----- 928
QY 1123 CTCATGCATCTCTGGGAGGTGGTGAGCCACATCGGCTGCTGCTGCTCAGGGAGA 1182
|||||:|||||
Db 929 -----ProCysProCys-----ProGluGlyPr 936
QY 1183 AGAGGCCGAGAAAGAACTGCTCCAAGTGCCTGGCAGCAGAGATGATCTGGAGGAAT 1242
|||||:|||||
Db 936 oGlySerGlnArgHisPheAlaThrSerCysHisArgAsp----- 949
QY 1243 TGAAGGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGGATTGAAATGGA 1302
```

```
Db 950 -GlyTyrSerGlnGlnIleValCysHisCysArgAlaGlyTyrThrGlyLeuArgCysG 969
|||||:|||||
QY 1303 GTAGAGTATAAG-----TG 1317
Db 969 uLaCysAlaProGlyHisPheGlyAspProSerLysProGlyGlyArgCysGlnLeuCy 989
QY 1318 TGAATGCATGTTGATTTTCTCTAGTCTAGAAATCTCTAGTTTAGAAAGGATGTTTAGG 1377
|||||:|||||
Db 989 sGluCysSer-----Gln 993
QY 1378 GAACATGAGGCTGGCTCTCAGCAACAAACCAGGCTCCCTCGCATCCCTGGGCCA----- 1432
|||||:|||||
Db 993 yAsnIle-----AspProThrAspProGlyAlaCysAspProHisTh 1007
QY 1433 -GGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCTGAGGAGG 1491
|||||:|||||
Db 1007 rGlyGlnCysLeuArgCysLeuHisHisThrGluGlyPro----- 1020
QY 1492 TGCATGGCCTGAGCATTTGTTGCTGGCCAGAGGAGAGAGCTTGGTTCCCATAGTCCT 1551
|||||:|||||
Db 1021 -----HisCys----- 1022
QY 1552 GGGAGAGTCTCTGCGAGGCGGCGAGGCGAGAGC-----AGCCCTCGGAGAGCTCAC 1605
|||||:|||||
Db 1023 -GlyHisCysLysProGlyPheHisGlyGlnAlaAlaArgGlnSerCysHisArgCysTh 1042
QY 1606 T-----CTGGTCGACTCTCTCTCAGAGAATGTGTCTGAGGAGCTCTCTGCATGAA 1659
|||||:|||||
Db 1042 rCysAsnLeuLeuGlyThrAspProGlnArg-----CysProse 1055
QY 1660 AACCTTAATGTTCTTCTGT 1678
Db 1055 rThrAspLeuCysHisCys 1061
```

RESULT 12

T42215
zonadhesin - mouse
N:Alternate names: sperm-specific membrane protein
C:Species: Mus musculus (house mouse)
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T42215
R:Gao, Z.; Garbers, D.L.
J. Biol. Chem. 273, 3415-3421, 1998
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane
A:Reference number: 222080; MUID:98123114; PMID:9452463
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-5376 <GAO>
A:Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1
C:Genetics:
A:Gene: Zan
A:Map position: 5
C:Function:
A:Description: functions in multiple cell adhesion processes
A:Note: found exclusively on the apical region of the sperm head
C:Keywords: cell adhesion

Alignment Scores:
Pred. No.: 0.151 Length: 5376
Score: 118.50 Matches: 137
Percent Similarity: 29.39% Conservative: 64
Best Local Similarity: 20.03% Mismatches: 232
Query Match: 3.10% Indels: 251
DB: 2 Gaps: 41

US-09-743-237-3 (1-2134) x T42215 (1-5376)

QY 16 CGAAGCTCCGGGGGGCGGCGAGCAGCGCGGAGCTCTCGGGGAGTACCCTGGGATCC 75

Db 3478 LysAspAlaGlnGlyAlaLeuLeuProAlaGlyLysThrTrpThrSerProGlyCysThr 3497

A:Accession: A54148
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2406 <LEV>
 C:Genetics:
 A:Gene: FlyBase:Ten-m
 C:Cross-references: FlyBase:FBgn000449
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:514-540/Domain: EGF homology <EGF>
 F:610-637/Domain: EGF homology <EGF1>

Alignment Scores:
 Pred. No.: 0.163 Length: 2406
 Score: 118.00 Matches: 132
 Percent Similarity: 31.7% Conservative: 79
 Best Local Similarity: 19.8% Mismatches: 240
 Query Match: 3.09% Indels: 214
 DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x A54148 (1-2406)

```

QY 7 GGTCAAGCGGAGCTCGCGGGGGGACAC---GCCAGCGGGGGAGCTCCTCGGGAGT 63
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GlySerGlyGlySerSerAlaAlaThrValThrAlaThrSerAsnSerGlyThr 335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 ACCCGGGATCCAGAGCTCAGCGCGCTGGAGGAGCTCGCGTCTCGCAGGCCCGCAGC 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 AlaGlnGlyLeuGlnSer-----ThrSerAlaSerAlaGluAlaThrSer 350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 CGCCCGCTGCGACAGTGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 351 -----SerAlaAlaThrSer-----SerSerGlnSer 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 TGTCTTTCCTCGCGCGCTCGGTCTCGGAGGAGCTCCACCGCGGGCTCGCCATG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 SerLeuThrProSerLeuSerSerSerLeuAlaAsnAlaAsnGlyGlyAlaArgThr 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 ATCCCAAGTGAATCAAGTAAGTACAGCTGCTACTACTACAAAGTAATAATCCGGAAGAAGC 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 PheProAla-ArgSerPheProAspGlyThrPheGly-----GlnI 395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 ACTTTGCGAGATCTCTCTGCTCAGAAATCTGTTCCAAATGTCCTCCAGTCCCGAGACT 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 395 eThrLeuGlyGlnLysLeuThrLysGluIleGlnProTyrSerTyrTrpAsnMetGlnPh 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 AGAGGATGCTCTCTGCTGTTCTCTTAAGAAAGAT----- 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 eTyrGlnSerGluProAlaTyrValLysPheAspTyrThrIleProArgGlyAlaSerI 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 -----TCCAAACCAATGGTGATATGCCAATTTGAAAGGGGCGACACAAAT 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 eGlyValTyrGlyArgArgAsnAlaLeuProThrHisThrGlnTyrHisPheLysGluVa 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 442 GCTATCTATAGACAATCTTAGACAGAGAACTAAAGCACTCCATTTG----- 490
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 lLeuSerGlyPheSerAlaSerThrArgThrAlaArgAlaAlaHisLeuSerIleThrAr 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 491 ----GTTCTCTAGTACAGATCAAAATAATATTATCTACAGTCA----- 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 gGluValThrArgTyrMetGluProGlyHisTrpPheValSerLeuTyrAsnAspAspG 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 -GATGCTCTAAACCAATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAAT 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 yAspValGlnGlu-----LeuThrPheTy 503
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 AATCTCATTAACACAACTTAGGGAGCCTTACCATCGGTAGTCAAGGGTCTCTCTTT 648
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 rAlaAlaValAlaGluAspMetThrGlnAsnCysProAsnGlyCysSerGlyAsnGly-- 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 CCCTCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGACTG 708
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 -----GlnCysLeuLeuGlyHisCysGlnCy 531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 709 CTTTGGC-----AGTGGGAGCTTTTGCACAACTGCAATTTGTAATAATTGTTGCAACA 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 531 sasnProGlyPheGlyAspAspCysSerGluSerValCysProValLeuCysSerG 551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 763 CTTGCATCATGATATTTAAACGGTTTAAAGCCATTAAGGCAATGCTTGGTAGAAATCCAGA 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 n---HisGlyGluTyrThrAsnGlyGlyCysIle-----Cys-----AsnProG 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 823 AGCTTTCCAGCCCAAAATTTGGGAGGGCCAAATTTGGCAATGTCAGCCCAAGCCAGCACAA 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 y-----TriPlysGlyLysGluCysSerLeuArgHisAspGluCysG 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 883 AGGGTGCAACTGCAGGAGGTGACGC---TGCCCTGAAGAATTAATCTCGAGTGC----- 931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 uValAlaAspCysSerGlyHisGlyHisCysValSerGlyLysCysGlnCysMetArgG 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 -TATGAGGCCCAAAATTTATGTTCTTCTATT---TGC-----AAATGCAATGCT-- 976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 yTyrLysGlyLys---PheCysGluGluValAspCysProHisProAsnCysSerGlyHI 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 977 -----TGCAAAATTTATGAAGAAGCCCAAGCCAGAACAA 1008
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 sGlyPheCysAlaAspGlyThrCysIleCysLysLysGlyTrpLysGlyProAspCysAl 638
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1009 GACA-----CTAATGAGCATGCCAAACTACATGCACACTGGAGTTTGA 1053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 638 aThrMetAspGlnAspAlaLeuGlnCysLeuProAsp----- 650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1054 AGCAGCCATTACCTGCCACCAAGAAATTTTCAGGACTTCCAAAGATTCAGTACAGATAG 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 651 -----CysSerGlyHisGlyThrPheAspLeuAspTh 661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1114 GCGGCTTCTCATGCATCTCCTGGGAGGTGGTGGAGGCCACATCGCCTGCTGCTTGC 1173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 rGln-----ThrCysThrCysGluAlaLy 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1174 T---CAGGGAGAGAGCGCGAGAAAGAACACATGC-----TCCAA 1209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 669 sTrpSerGlyAspAspCysSerLysGluLeuCysAspLeuAspCysGlyGlnHisGlyAr 689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1210 GTGCCCTGGCAGACAGATGATCTGGAG---GAATTTTGA-----AGTGCTTATCACA 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 689 gCysGluGlyAspAlaCysAlaCysAspProGluTrpGlyGlyGluTyrCysAsnThrAr 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 GATCTCCACACTGAGCTTTAAATCTAAGGATTTGAAATGGAGTACAGATATAAGTGTGA 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 gLeuCysAspValArgCysAsnGluHisGly-----GlnCysLy 722
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 A-----TGATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGATGT 1371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 722 sAsnGlyThrCysLeuCysValThrGlyTrpAsn----- 733
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1372 TTAGGGACATGAGGCTGCTCGACCAACACACAGGCTCCCTCGCATCCTCGGGCCC 1431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 734 -----GlyLysHisCysThrIleGluGlyCysProAsnSerCysAlaGlyHisGly-- 750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1432 AGGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTCTTCTGAGGAG 1491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 -----GlnCysArgValSerGlyGluG 758
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1492 TCATGCGCTGAGCATTTGTTGCTGGCCAGAGAGAGAGCTTGGTTTCCCATGCTCT 1551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 758 yGlnTrpGluCysArgCysTyrGluGly-----TipAspGlyProAspCy 773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1552 GGGAGAGTCTGTCAGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 773 sGlyIleAlaLeuGluLeuAsnCysGlyAspSerLysAspAsnAspLysAspGlyLeuVa 793
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1612 CGACTCTTCTCTCAGAGAAATGTTGCTCTGGAGGCTGCTGCTGCAAAACCCCTAATGTT 1671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 lAspCysGluAspPro-GluCysCysAlaSerHis----- 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1672 TTCTTGTGTTTTCATAATTTAGAAATAGTTTCTCCGGATGGGCTGTTGTGTATACC 1731
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 805 -----ValCysLysThrSerGlnLeuCysValSerAlaPro----- 816
 QY 1732 ACTTAAATCTCTAGAGACTACTGAACACCTCAAGATTTTCTCTAGCGTAGATATTTC 1791
 Db 817 -----LysProIleAspValLeuLeuArgLysGlnProp 828
 QY 1792 CCAGAGACAGC 1802
 Db 828 roAlaIleThr 831
 RESULT 14
 S47008
 tenascin-like protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C>Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S47008
 R:Baumgartner, S.; Martin, D.; Hagios, C.; Chiquet-Ehrismann, R.
 EMBO J. 13, 3728-3740, 1994
 A:Title: Ten(m), a *Drosophila* gene related to tenascin, is a new pair-rule gene.
 A:Reference number: S47008; MUID:94349920; PMID:8070401
 A:Accession: S47008
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2515 <BAU>
 A:Cross-references: EMBL:X73154; NID:g510505; PIDN:CAA51678.1; PID:g510506
 C:Genetics:
 A:Gene: FlyBase:Ten-m
 A:Cross-references: FlyBase:Fgn0004449
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:298-324/Domain: EGF homology <EGF>
 F:394-421/Domain: EGF homology <EGF1>
 Alignment Scores:
 Pred. No.: 0.163 Length: 2515
 Score: 116.00 Matches: 132
 Percent Similarity: 31.73% Conservative: 79
 Best Local Similarity: 19.85% Mismatches: 240
 Query Match: 3.09% Indels: 214
 DB: 2 Gaps: 35
 US-09-743-237-3 (1-2134) x S47008 (1-2515)
 QY 7 GGGTCAAGCGAAGCTCGCGGGGGCGACA---CGCAGCGCGGGAGCTCTCGGGAGT 63
 Db 100 GlySerGlyGlySerSerAlaAlaThrValThrAlaThrSerAsnSerGlyThr 119
 QY 64 ACCCGGGATCCAGAGCTCAGCGCGTGGAGAGCTGCGCTCTCGAGGCCCGCGAGC 123
 Db 120 AlaGlnGlyLeuGlnSer-----ThrSerAlaSerAlaGluAlaThrSer 134
 QY 124 CGCCCGCTGCAAGCTGCACCTTCCTGCTCTACCGCGCAGCGCCGCGGG 183
 Db 135 -----SerAlaAlaThrSer-----SerSerGlnSer 143
 QY 184 TGTTCCTCCCTCGGGCGCTGCTGCTCGAAGAGCCTCCACCGCGCGCTCCGCGATG 243
 Db 144 SerLeuThrProSerLeuSerSerSerLeuAlaAlaAsnAsnGlyGlyAlaArgThr 163
 QY 244 ATCCAGTTGNAATCAGGTAAAGCAGGTGGTACTACTACAGTAAATATCCGGNAGAGC 303
 Db 164 PheProAla-ArgSerPheProAspGlyThrThrPheGly-----GlnIle 179
 QY 304 AACTTTGCAGAAATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCTCAGTCCAGGAACT 363
 Db 179 eThrLeuGlnLysLeuThrLysGluIleGlnProTyrSerTyrTrpAsnMetGlnPh 199
 QY 364 AGAGGATCGCTCGCTGTTCTCTTAAGAAGAT----- 397
 Db 199 eTyrGlnSerGluProAlaTyrValLysPheAspTyrThrIleProArgGlyAlaSerIle 219
 QY 398 -----TCCACCAATGTTGATGCAATGCAATGAAAGGGGCGCACAAAT 441
 Db 506 sasnglythrCysLeuCysValThrGlyTrpAsn----- 517

Db 219 eGlyValTyrGlyArgArgAsnAlaLeuProThrHisThrGlnTyrHisPheLysGluVa 239
 QY 442 GCTATGTATAGACAATTTCTAGACAAGAGAACTAAAGACACTCCCATTTG----- 490
 Db 239 lLeuSerGlyPheSerAlaSerThrArgThrAlaArgAlaAlaHisLeuSerIleThrAr 259
 QY 491 -----GTTCTCAGTATCAAGATCAAAATAATATATCTACAGTCA----- 529
 Db 259 gGluValThrArgTyrMetGluProGlyHisTrpPheValSerLeuTyrAsnAspSpGl 279
 QY 530 -GATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTGCACCATCACAAAAT 588
 Db 279 yaspValGlnGlu-----LeuThrPheTy 287
 QY 589 AAATCTCATTAACAACAACACTTGAGGAGCGCTTACCATCGGTAGTCAACGGGTGCTTT 648
 Db 287 rAlaAlaValAlaGluAspMetThrGlnAsnCysProAsnGlyCysSerGlyAsnGly-- 306
 QY 649 CCCTCGGGATCAACTCTTCAGGACCACCACAAATAAATTTGGCTGGTACTGTGACTG 708
 Db 307 -----GlnCysLeuLeuGlyHisCysGlnCy 315
 QY 709 CTTTGGC-----AGTGGGACTTTTGCACAACACTGCAATTTGTAATTAATTTGTCACAA 762
 Db 315 sasnProGlyPheGlyGlyAspCysSerGluSerValCysProValLeuCysSerGl 335
 QY 763 CTTGCATCATGATTAACGCGTTTAAAGCCATTAAGGCATGCTTGGTGAATCCAGA 822
 Db 335 n---HisGlyGluTyrThrAsnGlyGluCysIle-----Cys-----AsnProGl 349
 QY 823 AGCTTTCCAGCCCAAAATTTGGGAAGGCCCAATTTGGCAATTTCAAGCCCCCAGCACAA 882
 Db 349 y-----TrpLysGlyLysGluCysSerLeuArgHisAspGluCysGl 363
 QY 883 AGGTCGCACTGCAGGAGTCAGGC---TGCCTGAACAATTTACTGCCAGTGC----- 931
 Db 363 uValAlaAspCysSerGlyHisGlyHisCysValSerGlyLysCysGlnCysMetArgGl 383
 QY 932 -TATGAGGCCCAATTTATGTTCTTCTTATT---TGC-----AAATGCTATTGT-- 976
 Db 383 yTyrLysGlyLys---PheCysGluGluValAspCysProHisProAsnCysSerGlyHi 402
 QY 977 -----TGCAAAATTTATGAAGAACGCCAGCACGAAA 1008
 Db 402 sGlyPheCysAlaAspGlyThrCysIleCysLysLysGlyTrpLysGlyProAspCysAl 422
 QY 1009 GACA-----CTAATGAGCATGCCAAATACATACGACAGTGGAGTTTGA 1053
 Db 422 aThrMetAspGlnAspAlaLeuGlnCysLeuProAsp----- 434
 QY 1054 AGCAGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAGATTCAGTCACGATAG 1113
 Db 435 -----CysSerGlyHisGlyThrPheAspLeuAspTh 445
 QY 1114 GCGGCTTCCTCATGCTCTCTGGGAGGTGGTGGAGCCACATGCGCTGCTGCTGTC 1173
 Db 445 rGln-----ThrCysThrCysGluAlaLy 453
 QY 1174 T---CAGGAGAACAGGCCCGCAGAAAGAACACTGC-----TCCAA 1209
 Db 453 sTrpSerGlyAspAspCysSerLysGluLeuCysAspLeuAspCysGlyGlnHisGlyAr 473
 QY 1210 GTGCTGGCAGACAGATGATCTCTGGAG---GAATTTTGA-----AGGTCTTATCACA 1260
 Db 473 gCysGluGlyAspAlaCysAlaCysAspProGluTrpGlyGlyClyTrpCysAsnThrAr 493
 QY 1261 GATTCTCCACACTGAGTTTAAATCTAAGGATTTGAAATCGAGTAGAGTATAAGTGTGA 1320
 Db 493 gLeuCysAspValArgCysAsnGluHisGly-----GlnCysLy 506
 QY 1321 A-----TGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGGATGT 1371
 Db 506 sasnglythrCysLeuCysValThrGlyTrpAsn----- 517

QY 1372 TTAGGGAAACATGAGGCTGGCTCTGCAGCAACAAACAGGCTCCCTCGCATCCCTGGGCC 1431
 Db 518 -----GlyLysHisCysThrIleGluLysCysProAsnSerCysAlaGlyHisGly-- 534
 QY 1432 AGGAGTCTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCTGAGGAGG 1491
 Db 535 -----GlnCysArgValSerGlyGluG 542
 QY 1492 TGCATGGCCCTGACCATTTGTTGCTGGCCAGAGGAGAGCTTGGTTCCTCATAGTCCT 1551
 Db 542 yGlnTrpGluCysArgCysTyrGluGly-----TrpAspGlyProAspCys 557
 QY 1552 GGAGAGTGTCTGCAGGGGGGGGAGGAGCAGAGCAGGCTCGGGAGAGCTCACTCTGTT 1611
 Db 557 sGlyIleAlaLeuGluLeuAsnCysGlyAspSerLysAspAsnAspLysAspGlyLeuVa 577
 QY 1612 CGACTCTTCTCAGAGAAATGTTCTGAGGCTGCTGCTGCATGAAACCCCTAATGTT 1671
 Db 577 lAspCysGluAspPro-GluCysCysAlaSerHis----- 588
 QY 1672 TTCTCTTTGTTTTCAAATATTAGAAATAGTTCTCGGATGGCTGTTGTGATACC 1731
 Db 589 -----ValCysLysThrSerGlnLeuCysValSerAlaPro----- 600
 QY 1732 ACTTAAATCTCTAGAGAACTACTGAACACACCTAAAGATTTCTGTAGCGTAGATATTCC 1791
 Db 601 -----LysProIleaspValLeuLeuArgLysGlnProp 612
 QY 1792 CCAGAGACACG 1802
 Db 612 roAlaIleThr 615

RESULT 15

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34513

R:Pavello, A.; Vaudin, M.

Submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.181	117.50	3507	117
Percent Similarity:	28.57%	Conservative:	67
Best Local Similarity:	18.17%	Mismatches:	170
Query Match:	3.07%	Indels:	290
DB:	2	Gaps:	33

US-09-743-237-3 (1-2134) x T34513 (1-3507)

QY 8 GGTCAAGCGAAGCTCGCGGGGGGCGACAGCGCGGGGAGCTCCTCGGGAGTACC 67
 Db 1268 GlySerGlyGluAla-----GlyValPro 1275
 QY 68 CGGGATCCAGAGCTCAGCGCGCTGGAGGACGTGGCGCTCCCTGCAGGC-----CCGCA 121
 Db 1276 GluSerProAspGlySerGlyGluSerThrSerAlaProAspGlyValSerProThr 1295

QY 122 GCGCCCGCTGCAACGTCACATTCCTGCTGCTGCTACCGCGCACCG----- 172
 Db 1296 SerSerAlaThrAlaProGluValProThrThrSerAlaSerThrProAspAlaVal 1315
 QY 173 CAGCCCGCGGGTGT-----TTGCCCTCGGGCGCCCTGGGT 208
 Db 1316 GluGluSerGlyIleProSerThrSerLysProThrAlaGluProLeuGluThrAla 1335
 QY 209 CCTGCCAAGAG-----CCTCCC 226
 Db 1336 ProSerThrGluValThrSerProGluGlySerGlyThrGluGluSerThrLeuProPro 1355
 QY 227 ACCCGGCGCTCGCATGATCCAGTTCAGTAAAGTAAGGTAAGCAGAGTGGTACTACTACAAGT 286
 Db 1356 ThrGluGlySer-----GlyGluSerThrThrSer 1365
 QY 287 ATAATCCG-----GAAGAAGCAACTTTGTCAGAAATCTTCTGCTCAGGAATCCTGTGC 340
 Db 1366 SerAlaProThrValGluProAlaThrVal----- 1375
 QY 341 AAGTTCCTCCATGGTCCCGAGGAGTCCCTGCTGCTGCTCTTAAAGAAAGATTCC 400
 Db 1375 ----- 1375
 QY 401 AACCCCAATGGTATATGCCAATTTGAAAGGGGCGACACAATGCTATATAGACAATCT 460
 Db 1375 ----- 1375
 QY 461 AGAACAAGAGAACTAAAGCACTCCATTTGGTTTCTCAGTATCAAGATCAAAATAATTAT 520
 Db 1376 -----LeuProGlnAsnArgAsnGluLys----- 1384
 QY 521 CTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTTCAGCAGCATCA 580
 Db 1385 -----ProGluPro 1387
 QY 581 ACAAAATAAATCTATTACACAACACTTGAGGGAGCCTTACCATCGGTAGTCAACGGG 640
 Db 1388 ThrLysAspThrPhe-----AlaLeuProThrThrThrGly 1400
 QY 641 TCTGCTTTCCCTCGGGATCAACTTT----- 667
 Db 1401 AlaProGlnAlaAsnAspSerSerValGluAsnThrLysCysThrSerSerAspGluCys 1420
 QY 668 -----CCAGCA----- 673
 Db 1421 GlyLeuAspAlaLeuCysGluArgArgThrGlyValCysArgCysGluProGlyPheGlu 1440
 QY 674 -----CCACCAAAATAAATTTGGCTGGTACTGTGACTGCTGCTTCCAGTGGGACTTT 727
 Db 1441 GlyAlaProProLysLysSerCysValAspValAspGluCys---AlaThrGlyAsp--- 1458
 QY 728 TGCACAACATGCNAATTTGTAATAATTGTCACAACACTTGCATCATGATATTGAACGGTTT 787
 Db 1459 ---HisAsnCysHisGluSerAlaArgCysGlnAsn-----Tyr 1470
 QY 788 AAAGCCATTAAAGCATGCTTGTGTAGAAATCCAGAGCTTCCAGCCAAATAATGGGAAG 847
 Db 1471 ValGlyGlyTyrAlaCysPhe-----CysProThrGlyPheArg---LysAlaAspAsp 1487
 QY 848 GGCCAATTGGCAATGCTCAAG-----CCCCAGCACACAAGGGTGCAACTGCAGGAGG 901
 Db 1488 GlySerCysGlnAspIleaspGluCysThrGluHisAsnSerThrCys----- 1503
 QY 902 TCAGGCTGCTGAAGAATTACTGCGAGTGCTATGAGGCCCAAAATATGTTCTTCTATT 961
 Db 1504 -----CysGlyAlaAsn 1507
 QY 962 TGCAAATGCTTGGT-----TGCAAAATATATGAGAAGC 997
 Db 1508 AlaLysCysValAsnLysProGlyThrThrSerCysGluCysGluAsn----- 1523
 QY 998 CCAGAACGAAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGC 1057

```
Db 1524 -----GlyPheLeuGly 1527
Qy 1058 AGCCATTAC-----CTGCCACCAAGAAATTTTCAGGACTTCCAAGATTTCAGTCACGAT 1111
Db 1528 AspGlyTyrGlnCysValProThrThrLys-----LysProCysAspSerThrGln 1544
Qy 1112 AGCGCGCCTTCCTCATGCATCTCCTGGAGGTGTGGAGGCACATGCGCCTGCTGCTT 1171
Db 1545 SerSerLysSerHisCys-----SerGluSerAsnMetSerCysGluVal 1559
Qy 1172 GCTCAGGAGAGAGCGGAGAGAACACACTGCTCCAAAGTCCTGCCA-----1219
Db 1560 AspThrValAspGlySerValGlu-----CysLysGluCysMetGlyGlyTyrLysLys 1577
Qy 1220 GAGCAGATGATCCTGGAGGATTTGGAAGTGTCTATCAGATTCCTCCACACTGAGTTT 1279
Db 1578 SerGlyLysValCysGluAspIleAsnGluCysValAlaGluLysAlaProCysSerLeu 1597
Qy 1280 AAATCTAAGGGATTGAAATGGAGTAGAGTATAAGTGTGAATGCATGTTGATTTGTCT 1339
Db 1598 AsnAlaAsnCysValAsnMetAsnGlyThrPheSerCysSerCys-----1612
Qy 1340 TAGTCTAGAAATCTAGTTTAAAGGATGTTTAGGGAACATGAGCTGCTGCTGCGAG 1399
Db 1613 -----LysGlnGly-TyrArgGlyAsp---GlyPheMetCysTh 1624
Qy 1400 C-----AACACACAGGCTCCCTGCTGCATCCCTGGGCCAGGAGTTTACTC 1444
Db 1624 rasPIleAsnGluCysAspGluArgHisProCysHisProHisAlaGluCysThrAsnLe 1644
Qy 1445 AGAGCTCTCTGAAGATGTGGCAACCCATGCCCTCCCTTTTCT-----1484
Db 1644 uGluGlySerPheLysCysGluCysHisSerGlyPheGluGlyAspGlyIleLysLysCy 1664
Qy 1485 -----GAGGAGGTGCATGGCCTG-----AG 1504
Db 1664 sThrAsnProLeuGluArgSerCysGluAspValGluLysPheCysGlyArgValAspHi 1684
Qy 1505 CATTGTTTGTCTGGCCAGAGAGAGAGCTTGGGTTCCCATAGTCTGGAGAGTGTCTG 1564
Db 1684 sValSerCysLeuSerValArgIleTyrAsnGly-----SerLeuSerSerValCy 1701
Qy 1565 CAGGCGCGGAGGAGAGAGAGGCGCTGCGGAGAGCTACTCTGTCGACTTCTCTCT 1624
Db 1701 sGluCysGluProGlyPheArgPheGluLysGluSerAsn-----SerCysValAspIl 1719
Qy 1625 CAGAGATGT 1634
Db 1719 eAspGluCys 1722
```

Search completed: April 21, 2003, 12:03:50
Job time : 118.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:50:04 ; Search time 30.5 Seconds
(without alignments)
5803.966 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 3824
Sequence: 1 aattcgggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=rlp
-Q/cgn2_1/USPTO.spool/US09743237/runat_21042003_112209_1461/app_query.fasta_1.2311
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237_CGN_1_1_47 -runat_21042003_112209_1461 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	43.1	299	1 MTL5_HUMAN	Q9v415 homo sapien
2	1251	32.7	295	1 MTL5_MOUSE	Q9wtj6 mus musculu
3	130.5	3.4	3695	1 LMA5_HUMAN	Q15230 homo sapien
4	123	3.2	1700	1 BAR3_CHITE	Q03376 chironomus
5	119.5	3.1	1172	1 AHM2_ARATH	O64474 arabidopsis
6	119.5	3.1	3010	1 POLG_HCVJA	P26662 h genome po
7	119	3.1	1576	1 YLK3_CAEEL	P41951 caenorhabdi
8	118.5	3.1	1801	1 LMB2_RAT	P15800 rattus norv
9	118.5	3.1	5376	1 ZAN_MOUSE	O88799 mus musculu
10	117	3.1	3011	1 POLG_HCV1	P26664 h genome po
11	117	3.1	3718	1 LMA5_MOUSE	Q61001 mus musculu
12	116.5	3.0	2003	1 NTC4_HUMAN	Q99466 homo sapien
13	116	3.0	3010	1 POLG_HCVBK	P26663 h genome po
14	116	3.0	3010	1 POLG_HCVJT	Q00289 h genome po
15	115.5	3.0	769	1 ITB2_HUMAN	P05107 homo sapien
16	114.5	3.0	778	1 TAST_HUMAN	Q12815 homo sapien
17	114.5	3.0	3726	1 TRX_DROME	P20659 drosophila
18	112.5	2.9	3133	1 HMCT_BOMMO	P98092 bombyx mori

ALIGNMENTS

RESULT 1

ID	MTL5_HUMAN	STANDARD;	PRT;	299 AA.
AC	Q9v415;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific			
DE	metalothionein-like protein).			
GN	MTL5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE-99208669; PubMed-10191092;			
RA	Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;			
RT	"A novel testis-specific metallothionein-like protein, tesmin, is an			
RT	early marker of male germ cell differentiation."			
RL	Genomics 57:130-136(1999).			
CC	-!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U86074; RAD24668.1; -			
DR	Genew; HGNC:7446; MTL5.			
DR	MTM; 604374; -			
DR	InterPro; IPR005172; CXC.			
DR	Pfam; PF03638; CXC; 1.			
DR	Spermatogenesis.			
FT	DOMAIN 99 191 CYS-RICH.			
SQ	SEQUENCE 299 AA; 33110 MW; 8f679e140f969c11 CRC64;			

19	112	2.9	3010	1	POLG_HCVTW	P29846 h genome po
20	111.5	2.9	1021	1	CARA_MOUSE	P58660 mus musculu
21	111	2.9	798	1	ITB7_HUMAN	P26010 homo sapien
22	110	2.9	1587	1	LMG3_HUMAN	Q9y6n6 homo sapien
23	109.5	2.9	573	1	C114_MOUSE	P19467 mus musculu
c	24	109	2.9	977	1 MCR_TURGB	Q29131 tupai glis
25	109	2.9	2499	1	MPRI_BOVIN	P08169 bos taurus
26	108.5	2.8	592	1	DLI3_MOUSE	O88516 mus musculu
27	108.5	2.8	1877	1	PK5_MOUSE	O04592 mus musculu
c	28	108	5376	1	ZAN_MOUSE	O88799 mus musculu
29	107.5	2.8	2282	1	ZAN_RABIT	P57999 oryctolagus
30	107	2.8	559	1	ECM1_MOUSE	Q61508 mus musculu
31	107	2.8	677	1	SP87_DICDI	P54643 dictyosteli
32	106.5	2.8	1786	1	LMB1_HUMAN	P07942 homo sapien
33	106	2.8	884	1	ANDR_EULFC	O97776 eulemur ful
34	106	2.8	1429	1	LI12_CAEEL	P14585 caenorhabdi
35	105.5	2.8	589	1	DLI3_RAT	O88671 rattus norv
36	105.5	2.8	971	1	RECK_HUMAN	O95980 homo sapien
37	105	2.7	618	1	DLI3_HUMAN	O9ny17 homo sapien
38	105	2.7	643	1	CD93_RAT	O9et61 rattus norv
39	104.5	2.7	769	1	ITB2_PIG	P53714 sus scrofa
40	104.5	2.7	1790	1	LMB1_DROME	P11046 drosophila
c	41	104	2.7	981	1 MCR_RAT	P22199 rattus norv
42	104	2.7	1173	1	TSPI_XENLA	P35448 xenopus lae
43	104	2.7	1520	1	ABL_DROME	P00522 drosophila
44	104	2.7	1687	1	Z142_HUMAN	P52746 homo sapien
45	104	2.7	1786	1	LMB1_MOUSE	P02469 mus musculu

Alignment Scores:

```
Pred. No.: 9.84e-121 Length: 299
Score: 1647.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.07% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-3 (1-2134) x MTL5_HUMAN (1-299)
QY 407 ATGGTGATATGCAATTCAGAGGGGACACAAATGCTATGTATAGACAATCTAGACA 466
Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 587 TTAATCTCATTACACAACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnGlnLeuGlyAlaLeuProSerValAlaAsnGlySerAla 80
QY 647 TTCCTCGGGATCAACTCTTCCAGGACCAACAAATAAATCTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 707 TCGTTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTATATATTTGTGCAACACTG 766
Db 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnAsnCysCysAsnLeu 120
QY 767 CATCATGATATTGAACGGTTAAAGCCATTAAAGGCATGCTTGGTGTAGAAATCCAGAACT 826
Db 121 HisHisAspIleGluArgPheLysAlaLeuLysAlaCysLeuGlyArgAsnProGluAla 140
QY 827 TTCAGCCCAAAATTTGGAAGGCCAATTTGGGCAATGTCAGCCCAACCAACAAAGGG 886
Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
QY 887 TGCACACTCAGAGGTGAGGCTGCTGAAGATTAATCTCGAGTGTATGAGGCCCAATTT 946
Db 161 CysAsnCysArgSerGlyCysLeuLysAsnTyrCysGlyCysTyrGluAlaGlnIle 180
QY 947 ATGTCTTCTCTATTGCAATGCAATTTGTCGCAAAATATTGAAGAAAGCCCAAGCA 1006
Db 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
QY 1007 AAGACACTAATGAGCATGCCAACTACATGAGAGTGGAGTGTGGAGGAGGAGCAATTTAC 1066
Db 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
QY 1067 CTGCCACCAAGAAATTTTCAGGACTTCCAGATTCAGTACGATAGGCGGCTTCCTCA 1126
Db 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
QY 1127 TGCATCTCTGGAGGTGGAGGCCACATGCGCTCGCTTGTCTGCTCAGGAGAGAGAG 1186
Db 241 CysIleSerTrpGluValAlaGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
QY 1187 GCCGAGAAAGAACACTGCTCCAAGTCCCTGGCAGAGCAGATCATCTCTGGAGGAATTTGGA 1246
Db 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGly 280
QY 1247 AGGTCTTATCACAGATTCTCCACTGAGTTAAATCTAAGGATTTGAAATGGAG 1303
Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 2
MTL5_MOUSE
ID MTL5_MOUSE STANDARD: PRT: 295 AA.
```

```
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RC MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U77383; AAD24667.1; -
CC EMBL; U67176; AAD24666.1; -
CC MGD; MGI:1340029; MTL5.
CC InterPro; IPR005172; CXC.
CC Pfam; PF03638; CXC; 1.
CC KW Spermatogenesis.
CC FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 7.46e-90 Length: 295
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 32.71% Indels: 4
DB: 1 Gaps: 1

US-09-743-237-3 (1-2134) x MTL5_MOUSE (1-295)
QY 407 ATGGTGATATGCAATTCAGAGGGGACACAAATGCTATGTATAGACAATCTAGACA 466
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 587 TTAATCTCATTACACAACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAlaAspAsnGlyAlaLeuProSerAlaValAsnGlyAla 80
QY 647 TTCCTCGGGATCAACTCTTCCAGGACCAACAAATAAATCTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 707 TCGTTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTATATATTTGTGCAACACTG 766
```



```

CC EMBL; AL354836; CAC22309.1; ALT_SEQ.
DR EMBL; AL354836; CAC22310.1; -.
DR EMBL; AB067494; BAB67800.1; -.
DR EMBL; AB011105; BAA25459.1; -.
DR EMBL; Z95636; CAB09137.1; -.
DR HSP; P02468; IKLO.
DR Genew; HGNC:6485; LMA5.
DR MIM; 601033; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 2.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 35
FT CHAIN 36 3695
FT DOMAIN 36 299 LAMININ ALPHA-5 CHAIN.
FT DOMAIN 300 358 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 359 428 LAMININ EGF-LIKE 1.
FT DOMAIN 429 474 LAMININ EGF-LIKE 2.
FT DOMAIN 494 540 LAMININ EGF-LIKE 3.
FT DOMAIN 541 586 LAMININ EGF-LIKE 4.
FT DOMAIN 587 631 LAMININ EGF-LIKE 5.
FT DOMAIN 632 676 LAMININ EGF-LIKE 6.
FT DOMAIN 677 722 LAMININ EGF-LIKE 7.
FT DOMAIN 723 775 LAMININ EGF-LIKE 8.
FT DOMAIN 776 828 LAMININ EGF-LIKE 9.
FT DOMAIN 829 850 LAMININ EGF-LIKE 10.
FT DOMAIN 851 1437 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1638 1830 LAMININ EGF-LIKE 17.
FT DOMAIN 1831 1863 LAMININ EGF-LIKE 18.
FT DOMAIN 1864 1912 LAMININ EGF-LIKE 19.
FT DOMAIN 1913 1968 LAMININ EGF-LIKE 20.
FT DOMAIN 1969 2022 LAMININ EGF-LIKE 21.
FT DOMAIN 2023 2069 LAMININ EGF-LIKE 22.
FT DOMAIN 2070 2116 LAMININ EGF-LIKE 23.
FT DOMAIN 2117 2166 LAMININ EGF-LIKE 24.
FT DOMAIN 2167 2735 LAMININ II AND I.
FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
FT DOMAIN 3124 3292 LAMININ G-LIKE 3.
FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
FT DOMAIN 2203 2221 COILED COIL (POTENTIAL).
FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
FT DOMAIN 2510 2670 COILED COIL (POTENTIAL).
FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1838 1840 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 302 322 BY SIMILARITY.
FT DISULFID 324 333 BY SIMILARITY.
FT DISULFID 336 356 BY SIMILARITY.
FT DISULFID 359 368 BY SIMILARITY.
FT DISULFID 361 393 BY SIMILARITY.
FT DISULFID 396 405 BY SIMILARITY.
FT DISULFID 408 426 BY SIMILARITY.
FT DISULFID 429 440 BY SIMILARITY.
FT DISULFID 431 447 BY SIMILARITY.

```

```

FT DISULFID 449 458 BY SIMILARITY.
FT DISULFID 461 471 BY SIMILARITY.
FT DISULFID 494 506 BY SIMILARITY.
FT DISULFID 496 515 BY SIMILARITY.
FT DISULFID 517 526 BY SIMILARITY.
FT DISULFID 529 538 BY SIMILARITY.
FT DISULFID 541 553 BY SIMILARITY.
FT DISULFID 543 560 BY SIMILARITY.
FT DISULFID 562 571 BY SIMILARITY.
FT DISULFID 574 584 BY SIMILARITY.
FT DISULFID 587 599 BY SIMILARITY.
FT DISULFID 589 605 BY SIMILARITY.
FT DISULFID 607 616 BY SIMILARITY.
FT DISULFID 619 629 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT DISULFID 634 650 BY SIMILARITY.
FT DISULFID 652 661 BY SIMILARITY.
FT DISULFID 664 674 BY SIMILARITY.
FT DISULFID 677 689 BY SIMILARITY.
FT DISULFID 679 696 BY SIMILARITY.
FT DISULFID 698 707 BY SIMILARITY.
FT DISULFID 710 720 BY SIMILARITY.

Alignment Scores:
Pred. No.: 0.0298 Length: 3695
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
DB: 1 Gaps: 28

US-09-743-237-3 (1-2134) x LMA5_HUMAN (1-3695)
QY 18 AAGTCGCGGGGGCGGACAGCGCGGGGAG-----CTCCTCGGGGAGTAC 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 LysLeuValGlyGlyProValAlaGlyAspProAsnGlnThrIleGlnGlyGluTyr 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 CCGGGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleasp 122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 96 GAGCTC-----GCGCTCTCGAGCGCCGCGAGCGCGCGCTGC 134
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal 142
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 AACGTG-----CACTTCCTGCTCGCTCGCTACCCGCGCAC 170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 171 CGACGCCGCGGGGTGTTTGGCCCTGGCGGCGCTGCGGAGGAGCGCTCCACCC 230
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 231 GGGCGTCGCGATGATCCAGTTGAAATCAAGTAAGCAGGTGGTACTACTACAAGTAATA 290
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 GlyArgThrTyrGlnPro-----TrpGlnPhePhe 188
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 291 ATCCGGAAGAGCACTTTCGAGATCTTCTGCTCAGGAATCCTGTGCAAGTTCCCAT 350
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 351 GGTCCCGAGGAATA-----GAGGATGCTCCTGCTGCTCTTAAAGAAA 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaIleCysThrThrGluTyr 219
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 GATTCCAAACCAATGGTGATATGCCAATGAAGGGGGGACACAAATGCTATGATAGAC 454
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 455 AATCTAGACACAGAGCACTAAAGCAGCTCCATTGGTTCTCAGTACAGATCAAAAT 514
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 AsnGlyArgProGlyAlaMet----- 242
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


FT DOMAIN 338 345 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 346 363 POTENTIAL.
FT DOMAIN 364 656 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 657 676 POTENTIAL.
FT DOMAIN 677 680 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 681 700 POTENTIAL.
FT DOMAIN 701 1172 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1172 1172 HMA.
FT MOD_RES 1157 1169 POLY-HIS.
FT METAL 401 401 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 601 601 MAGNESIUM (BY SIMILARITY).
FT METAL 605 605 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1172 AA; 127208 MW; 1E913B9450443440 CRC64;

Alignment Scores:
Pred. No.: 0.17 Length: 1172
Score: 119.50 Matches: 110
Percent Similarity: 29.77% Conservative: 57
Best Local Similarity: 19.57% Mismatches: 184
Query Match: 3.12% Indels: 211
DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x AHW2_ARATH (1-1172)

QY 235 GTCCGCATGCCAGTTGAAATCAAGTAAAGCAAGTGGTACTACTACAAAGTAATAATCC 294
Db 639 IleArgArgIleProGlnAlaValLysLeuAlaArg-----Arg 651
QY 295 GGAAGAAGCAACTTTCAGAAATCTCTGCTCAGGNAATCCTGTTCGAAAGTCCCATGGTC 354
Db 652 AlaArgArgLysValValGluAsnValCysLeuSerIleLeuLys----- 667
QY 355 CCAGGAACATAGAGGATGCTCTGCTCTTAAAGAAAGATTCAACCCCAATGGTG-- 412
Db 668 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 582
QY 413 ----ATATGCCAATTTGAAGGGGGCACAATGCTATGATATAGCAATTCAGA----- 463
Db 682 pAlaAlaValLeuValAspValGlyThrCysLeuLeuValIlePheAsnSerMetLeuLe 702
QY 464 -ACAAGAGAACTAAAGCACTCCATTTGGTTCTCTAGTATCAAGATCAAAATAATTATCT 522
Db 702 uLeuArgGluLysLysLysIleGlyAsnLysLysCysTyrArg----- 716
QY 523 ACAGTCAGATGCTCTAAACCAATGACTCTTTAGTAGGAGATTTTCCAGCATCAAC 582
Db 717 -----AlaSerThrSe 720
QY 583 AAAATAAATCATTTACACAACTTGAGGAGCCTTACCATCGGTAGTCAAC----- 637
Db 720 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuGlu 738
QY 638 ----GGGCTGCTTCCCTCCGGATCACTCTTCCAGGACCACCAAAATAACTTTGGC 693
Db 738 uAlaGlyLeuLeuThrLysSerGlyAsn----- 747
QY 694 TGGGTACTGTACTGCTTTGCCAGTGGGACTTTTGCACAACTGCAAT----- 742
Db 748 -GlyGlnCysLysSerSerCysGlyAspLysLysAsnGlnGluAsnValValMetMe 767
QY 743 -----TGT----- 745
Db 767 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLys 787
QY 746 -----AATAATGTTTC----- 757
Db 787 sGluGluLysValLysProLeuValLysAspGlyCysSerGluLysThrArgLysSe 807
QY 758 -----AACAACTTCATCATATAT 777
Db 807 rGluGlyAspMetValSerLeuSerSerCysLysLysSerHisValLysHisAspLe 827
QY 778 TGAACGGTTTAAAGCATTAAGGCATGCTTGGTAGAAATCCAGAAGCTTTCCAGCCAAA 837

Db 827 u---LysMetLysGlyGlySerGlyCysAlaSerLysAsnGluLysGlyLysGluVa 846
QY 838 AATTGGGAAGGCCAATTGGGCAATGTCAGCCCAAGCACAACAAGGTTGCAACTGCAG 897
Db 846 lValAlaLysSerCysCysGluLysProLysGlnValGluSerValGlyAspCysL 866
QY 898 GAGGTGAGCTGCTGCTGAAG----- 916
Db 866 sSerGlyHisCysGluLysLysGlnAlaGluAspIleValProValGlnIlel 886
QY 917 ----AATTACTCGAGTGTATGAGGCCCAAAATTTGTTCTTCTATTTCGAAATGCAT 972
Db 886 eGlyHisAlaLeuThrHisValGluIleGluLeuGlnThrLysGluThrCysLysThrSe 906
QY 973 TGGTTGCCAAAATTTAAGAAAGCCCAAGAACAAAGACACTAATGAGCATGCCAAAC-- 1030
Db 906 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuLeuSerSerGluAsnTh 926
QY 1031 ----TACATGCAGACTGGAGTTTG-----GAAGGCAGCCA 1062
Db 926 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl 946
QY 1063 TTACCTGCCCAACCAAGAAATTTTCAGGACTTCCAAAGATTTCAGTACAGTACGAGCCTTC 1122
Db 946 uAsnMetGlyThrValLysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 966
QY 1123 CTCATGCATCTCTCGGAGGTGGTGAGGCCACATGCGCTGCTGCTGCTCAGGGAGA 1182
Db 966 rGlyGluIleThr-----LeuAlaSerGluGl 975
QY 1183 AGAGCGCGAAGAAAGCACTGCTCCAAAG-----TGCCTGGCAGAGCAGCATCTCTGA 1236
Db 975 uGluThrAspAspGlnAspCysSerSerGlyCysCysValAsnGluGlyThrValLysGl 995
QY 1237 GGAATTTGGAAGGTGCTTATCAGATTCTCCACTGATGTTTAAATCTAAGGATTTGAA 1296
Db 995 nSerPheAspGluLysLysHisSerValLeu-----ValGluLysGluGlyLeuAs 1012
QY 1297 AATGAGTAGACTATAAAGTGTGAATGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTA 1356
Db 1012 pMetGluThrGlyPheCysCysAspAlaLysLeuValCys----- 1025
QY 1357 GTTTAGAAAGGATGTTTAGGGGAACATAGGCTGCTGTCAGCAACAACAGGCTCCCC 1416
Db 1025 ----- 1025
QY 1417 TCATCTCCCTGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCC 1476
Db 1026 -----CysGlyAsnThr----- 1029
QY 1477 CTTTTCTGAGGAGGTGCTGCTGAGCATTGTT---TTGCTGCGCCAGAGAGAGAGC 1533
Db 1030 -----GluGlyGluValLysGluGlnCysArgLeuGluIleLysLysGluGlu-- 1045
QY 1534 TTGGGTTCCATAGTCTGGGAGGTGCTGAGGCGCGGAGGAGCAGCAGAGGCGCTG 1593
Db 1046 -----HisCysLysSerGlyCysGlyGluGluIleGlnTh 1058
QY 1594 CGGAGAGTCACTCTGCTGAGTCTTCTCTCAGAGAAATGT-----TGCTTGGAGGCT 1647
Db 1058 rGlyGluIleThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 1078
QY 1648 GC 1649
Db 1078 ys 1078
RESULT 6
POLG_HCVJA
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

Db 1284 ThrIleThrThrGlyGlyPro-----IleThrTyrSerThr 1295
QY 335 TGTTCAGAGTCCCATGGTCCAGCAACTAGAGGATGCTCTGCTGTCTCTTAAGAAA 394
Db 1296 TyrCysLysPhe-----LeuAlaAspGlyGlyCysSerGlyGlyAlaTyr 1310
QY 395 GATTCACACCCCAATGGTATGATGCAATTTGAAGGGGCACAAATGCTATGTATA--- 451
Db 1311 Asp-----IleIleIleCysAspGluCysHisSerThrAspSerThrThrIleLeu 1327
QY 452 -----GACAAATCTAGACAACAGAACTAAAGACACTCCATTTGGTT 493
Db 1328 GlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAla 1347
QY 494 -----CCTCAGTATCAAGATCAAAATAAT 517
Db 1348 ThrAlaThrProGlySerIleThrValProHisProAsnIleGluValAlaLeu 1367
QY 518 TATCTACAGTACAGTCCCT-----AAACAAATGACTGCTTTAGTAGG 562
Db 1368 SerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIleLysGly 1387
QY 563 AGA-----TTTTTGGCAGCATCAACAAATTAATCTCATACACAACTT 610
Db 1388 GlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeu 1407
QY 611 GAGGGA---GCCTTACCATCGGTA-----GTCAACGGGTCTGCTTTCCCC 652
Db 1408 ThrGlyLeuGlyLeuAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIlePro 1427
QY 653 ---TCGGGATCAACTCTCCAGGACACCAAAATAACTTTGGTGGGTACTGTGACTGC 709
Db 1428 ThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyPhe----- 1444
QY 710 TTTGCCAGTGGGACTTTTGCACAACTGCAATTTGTAATTTTGTGCAACAACTG--- 766
Db 1445 -----ThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrVal 1462
QY 767 -----CATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATTAAAGCCATT 1482
Db 1463 AspPheSerLeuAspProThrPheThrIleGluThrThrThrLeuProGlnAspAlaVal 1482
QY 809 GTAGAAATCCAGAGCTTTCCAGCCAAATAATTTGGAAAGGCCAATTTGGC----- 859
Db 1483 SerArgAlaGlnArg-----ArgGlyArgThrGlyArgGlyArgSerGlyIleTyrArg 1500
QY 860 ATGTCAAGCCCCACACCAACAAAGGTGCAACTGCAGGAGTGCAGGCTGCTGAAGAT 919
Db 1501 PheValThrProGlyGlu-----ArgProSerGlyMetPheAspSer 1514
QY 920 -----TACTGCGAGTGTATGAGGCCAAATTTATGTGTTCTTCTATTGCAAAATGCATT 973
Db 1515 SerValLeuCysGluCysTyrAspAla----- 1523
QY 974 GTTTCACAAAATTTATGAAGAACCCAGAACGAAACACTAATGAGCATGCCAAACTAC 1033
Db 1524 GlyCysAlaTyrTyrGluLeuThrPro---AlaGluThrSerValArgLeuArgAlaTyr 1542
QY 1034 ATGCAGACTGGAGGTTTGAA-----ACGATAGGCGGCTCTCTCATGCTCCCTGG 1054
Db 1543 LeuAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPhe 1562
QY 1055 -----GGCAGCATTAACCTGCGCCACCAACCAAAATTTTCAGGA----- 1090
Db 1563 ThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsn 1582
QY 1091 CTTTCAAGATTACGTC-----ACGATAGGCGGCTCTCTCATGCTCCCTGG 1138
Db 1583 LeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProPro 1602
QY 1139 GAGGTGG-----TGAGGCCACATCGCCCTGCTTGTCTCAGGAGAAAGAGCC 1189
Db 1602 oSerTrpAspGlnMetTrpLys-----CysLeuIleArgLeuLysProTh 1617

QY 1190 GAGAAAGAACAACACTGCTCAAGTGCTGGCAGAGCAGATGATCCTGGAGGAATTTGGAAGG 1249
Db 1617 rLeuHisGlyProThrProLeuLeuTyrArgLeu-----G1 1629
QY 1250 TGCTTATCACAGATCTCCACACTGAT----- 1277
Db 1629 yAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIleMetAlaCysMe 1649
QY 1278 -----TTAAATCTAAGGGATTGAATGGAGTAGAGTATAGTGAATGCAATGCATG 1327
Db 1649 tSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAl 1669
QY 1328 TTGATTTTGTCTTACTAGTAAATCTCTAGTTTGAAGAGGATGTTTAGGGGAACATGAGG 1387
Db 1669 aLeuAlaAlaTyrCysLeuThrThrGlySerValValIleVal-----G1 1684
QY 1388 CTGGCTCTGCAGCAACACAGGCTCCCTGCATCCCTGGGGCCCGAGGAGTGTACTACA 1447
Db 1684 yArgIleIleLeuSerGlyArgProAlaValIleProAspArgGluValLeuTyrGlnG1 1704
QY 1448 GCTCTCT-----GAAGATGTGGCAACCATCCCTTTCTTGAGGAGGTGCATGGCCT 1501
Db 1704 uPheAspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGln-----GlyMe 1722
QY 1502 GAGCATTTGTTCTGCGCCAGAGGAGAGCTTGGGTTCCCATAGTCTCTGGGAGAGTGT 1561
Db 1722 tClnLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-----LeuGlnThrAl 1739
QY 1562 CTGCAGGCGGCGGAGGAGGAGGAGGCGGCTCGGAGAGCTCACCTCTGGTGCACCTTT 1619
Db 1739 aThrLysGlnAlaGluAla---AlaAlaProValValGluSerLysTrpArgAlaLeu 1757
RESULT 7
YLK3_CABEEL STANDARD; PRT; 1576 AA.
AC P41951; 0950P7;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative serine/threonine-protein kinase D1044.3 in chromosome III
DE (EC 2.7.1.-).
GN D1044.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Pauley A.; Waterston R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RW [2]
RE REVISIONS, AND ALTERNATIVE SPLICING.
RA Waterston R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a (shown here) and b; are
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STRONG, TO ZC84.1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

CC EMBL; U00065; AAK68286.1; -;
CC EMBL; U00065; AAL27237.1; -;
CC HSSP; Q63450; 1A06.
CC WormPep; D1044.3a; CE27894.

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zonadhesin precursor.
 GN ZAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=98123114; PubMed=9452463;
 RA Gao Z., Garbers D.L.;
 RA "Species diversity in the structure of zonadhesin, a sperm-specific,
 RT membrane protein containing multiple cell adhesion molecule-like
 RT domains.";
 RT J. Biol. Chem. 273:3415-3421(1998).
 RL [2]
 RP SEQUENCE OF 4864-5376 FROM N.A.
 RN TISSUE=Testis;
 RX MEDLINE=97271566; PubMed=9126492;
 RA Gao Z., Harumi T., Garbers D.L.;
 RA "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD.
 CC -1- TISSUE SPECIFICITY: IN TESTIS: PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U97068; AAC26680.1;
 DR EMBL: U83190; AAC53125.1;
 DR MGD; MGI:106656; ZAN.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003645; FOLN.
 DR InterPro: IPR000998; MAM domain.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR003328; TILA_Cysrich.
 DR InterPro: IPR001007; VWF_C.
 DR InterPro: IPR001846; VWF_D.
 DR Pfam; PF00094; vwd; 4.
 DR Pfam; PF00629; MAM; 3.
 DR Pfam; PF01826; TIL; 25.
 DR Pfam; PF02345; TILA; 25.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00274; FOLN; 11.
 DR SMART; SM00137; MAM; 2.
 DR SMART; SM00214; VWC; 17.
 DR SMART; SM00216; VWD; 4.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 18.

DR PROSITE; PS00740; MAM_1; FALSE_NEG.
 DR PROSITE; PS50060; MAM_2; 3.
 KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
 KW Repeat.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 5376 ZONADHESIN.
 FT DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 5311 5337 POTENTIAL.
 FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 45 210 MAM 1.
 FT DOMAIN 215 374 MAM 2.
 FT DOMAIN 377 542 MAM 3.
 FT DOMAIN 547 1170 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
 FT (MUCIN-LIKE DOMAIN).
 FT DOMAIN 1171 1280 VWFD 1 (PARTIAL).
 FT DOMAIN 1281 1669 VWFD 2.
 FT DOMAIN 1670 2056 VWFD 3.
 FT DOMAIN 2057 2459 VWFD 4.
 FT DOMAIN 2460 2579 VWFD 5 (PARTIAL).
 FT DOMAIN 2580 2699 VWFD 6 (PARTIAL).
 FT DOMAIN 2700 2819 VWFD 7 (PARTIAL).
 FT DOMAIN 2820 2939 VWFD 8 (PARTIAL).
 FT DOMAIN 2940 3059 VWFD 9 (PARTIAL).
 FT DOMAIN 3060 3179 VWFD 10 (PARTIAL).
 FT DOMAIN 3180 3299 VWFD 11 (PARTIAL).
 FT DOMAIN 3300 3416 VWFD 12 (PARTIAL).
 FT DOMAIN 3417 3536 VWFD 13 (PARTIAL).
 FT DOMAIN 3537 3656 VWFD 14 (PARTIAL).
 FT DOMAIN 3657 3776 VWFD 15 (PARTIAL).
 FT DOMAIN 3777 3892 VWFD 16 (PARTIAL).
 FT DOMAIN 3893 4029 VWFD 17 (PARTIAL).
 FT DOMAIN 4029 4148 VWFD 18 (PARTIAL).
 FT DOMAIN 4149 4263 VWFD 19 (PARTIAL).
 FT DOMAIN 4264 4283 VWFD 20 (PARTIAL).
 FT DOMAIN 4384 4503 VWFD 21 (PARTIAL).
 FT DOMAIN 4504 4623 VWFD 22 (PARTIAL).
 FT DOMAIN 4624 4743 VWFD 23 (PARTIAL).
 FT DOMAIN 4744 4863 VWFD 24 (PARTIAL).
 FT DOMAIN 4864 5261 VWFD 25.
 FT DOMAIN 5259 5295 EGF-LIKE.
 FT DISULFID 5263 5274 BY SIMILARITY.
 FT DISULFID 5268 5283 BY SIMILARITY.
 FT DISULFID 5285 5294 BY SIMILARITY.
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1216 1216 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1314 1314 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1814 1814 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1933 1933 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2028 2028 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2111 2111 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2142 2142 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2332 2332 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2575 2575 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2692 2692 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2812 2812 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 5376 AA; 579908 MW; 0B44DB77DF2A2620 CRC64;

Alignment Scores:
 Pred. No.: 0.278 Length: 5376
 Score: 118.50 Matches: 137
 Percent Similarity: 29.39% Conservative: 64
 Best Local Similarity: 20.03% Mismatches: 232
 Query Match: 3.10% Indels: 251
 DB: 1 Gaps: 41

US-09-743-237-3 (1-2134) x ZAN_MOUSE (1-5376)

QY 16 CGAAGTCGGGGGGGCGACAGCGGGGGGAGCTCGGGGGAGTACCCCGGATCC 75
 Db 3478 LysAspAlaGlnGlyAlaLeuIleProAlaGlyLysThrTrpThrSerProGlyCysThr 3497
 QY 76 CAGAGTCAGCGCGCTGGAGGAGCTCGCGTCTCGAGGCCCGCGAGCCCGCCGCTGCA 135
 Db 3498 GlnSerCysAlaCysMetGlyGlyAlaValGlnCysGlnSerSerGlnCysProGly 3517
 QY 136 ACG---TGC-----ACTTCCTGCTCTCGC-----TGC 159
 Db 3518 ThrTyrcysLysAspGlnGluAspGlyAsnSerAsnCysAlaLysIleThrLeuGlnCys 3537
 QY 160 TACCCGCGCACCCGCGCGGTGTGTTGCCCTCGGGCGCCT-----204
 Db 3538 ProAlaHisSerLeuPheThrAsnCysLeuProCysLeuProSerCysLeuAspPro 3557
 QY 205 ---GGTCTCGCAAGAGGCTCCACCGCGGCTCGGATGATCCAGTTCAGATCAAG 261
 Db 3558 AspGlyLeuCysGlyAlaSer-----3565
 QY 262 GTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGAAGCACTTTCAGAACTCTCT 321
 Db 3565 -----3565
 QY 322 TGCTCAGGAATCCTGTTGCAAGTTCCTCGTCCCGAGAACT-----AGAGGATCCCTC 375
 Db 3566 -----ProLysValProSerThrCysLysGluGlyCysIle 3577
 QY 376 CTGC-----TGTTCTCTTAAG-----391
 Db 3578 -CysGlnSerGlyTyrrValLeuSerAsnAsnLysCysLeuLeuArgAsnArgCysGlyCy 3597
 QY 392 -AAAGATTCCAAACCAATGGTGATA-----391
 Db 3597 sLysAspAlaHisGlyAlaLeuIleProGluAspLysThrTrpValSerArgGlyCysTh 3617
 QY 416 -----TGCCAAATGAAAGGGGCACA---CAAATGCTA-----445
 Db 3617 rGlnSerCysValCysThrGlyGlySerIleGlnCysLeuSerSerGlnCysProProGl 3637
 QY 446 -----TGTATAGCAATCTAGACACAGAGAACTAAAGCACTCCATTTGGTCTCTCA 498
 Db 3637 yAlaTyrcysLysAspAsn---GluAspGlySerSerAsnCysAlaArgIleProProGl 3656
 QY 499 GTATCAAGATCAAAATATATCTACAGTCAGATGCTCCCTAAACCAATCACTGCTTTACT 558
 Db 3656 nCysProAlaAsnSerHisTyrrThrAspCysPheProProCysProProSerCysSerAs 3676
 QY 559 AGGAGATTTTGGCCAGCATCAACAAATAATATCTATTACACACAACTTGAAGGAGC 618
 Db 3676 pProGluGlyHisCysGluAlaSerGlyProArgValLeuSerThrCysArgGluGlyCy 3696
 QY 619 CTTA-----CCATCGGTA-----631
 Db 3696 sLeuCysAsnProGlyPheValLeuAspArgAspLysCysValProArgValGluCysGl 3716
 QY 632 -----GTCAAGGGTGTGCTTCCCTCGGGATCAACTCTTCCAGGACCA-----676
 Db 3716 yCysLysAspAlaGlnGlyAlaLeuIleProSerGlyLysThrTrpThrSerProGlyCy 3736

QY 677 -----CCAAAAATAACTTTGGCTGGGTACTGTGACTGTGCTTGGCCAGT-----718
 Db 3736 sThrGlnSerCysAlaCysMetGlyGlyValGlnCysGlnSerSerGlnCysProPr 3756
 QY 719 -GGGACTTTTGC-----AACAACTCAATTGTAATAATTGTTGGCAACAA 762
 Db 3756 oGlyThrTyrcysLysAspAsnGluAspGlyAsnSerAsnCysAlaLysIleThrLeuGl 3776
 QY 763 CITGCATCATGATATTGAACGGGTTTAAAGCCATTAAAGCATGCTCTTGGTAGAATCAGA 822
 Db 3776 nCysProThrHisSerAsnTyrrThrAspCysLeuProPheCysLeu-----Prose 3793
 QY 823 AGCTTTCCAGCCCAAAATTTGGGAAGGGCCAATTGGGCAATGTCAAGCCCCAGCAC-----877
 Db 3793 rCysLeuAspPro-----SerAlaLeuCysGlyGlyThrSerProLysGlyProse 3810
 QY 878 -----ACAAGAGGTGCAACTGCAGG-----AGTCAAGCTGC--910
 Db 3810 rThrCysLysGluGlyCysValCysGlnProGlyTyrrValLeuAspLysAspLysCysI 3830
 QY 911 -CTGAGAATTACTCGAGTGTCTATGAGCCCAA-----943
 Db 3830 eLeuLysIleGluCysGlyCysArgAspThrGlnGlyAlaValIleProAlaGlyLysTh 3850
 QY 944 -----ATTATGTGTTCTTCTTATTGCAAAATGC-----ATTGTTGCAA 981
 Db 3850 rTrpLeuSerThrGlyCysIleGlnSerCysAlaCysValGluGlyThrIleGlnCysGl 3870
 QY 982 AATTATGAAGAAAGCCAGAA-----CGAAAGACACT 1014
 Db 3870 nAsnPheGlnCysProProGlyThrTyrrCysAsnHisAsnAsnCysAlaLysIlePr 3890
 QY 1015 AATGAGCATGCAAACTACATCAGACTCGAGGTTTGAAGGAGCAGCATTAATCTGCCACC 1074
 Db 3890 oLeuGlnCysProAlaHis-----SerHisPheThr-----3900
 QY 1075 AACGAAATTTTCAGGACTTCCAAGATTCACTCAGATAGCGGCTTCTCATGATCTC 1134
 Db 3901 -----SerCysLeuProSerCys-----ProProSerCysAlaAs 3912
 QY 1135 CTGGAGGTGGTGGAGGCCACATGCGCTGCTGCTCAGGAGAGAGAGCCGAGAA 1194
 Db 3912 n-----LeuAspGlySerCys-----GluGlnThrSerProLysValPr 3925
 QY 1195 AGAACACTGCTCCAAG---TGCTGCGCAGCAGATGATCTCTGGAGGAATTTGGAGGTG 1251
 Db 3925 oSerThrCysLysGluGlyCysLeuCysGlnProGlyTyrrPheLeuAsnAsnGlyLysCy 3945
 QY 1252 CTATCACAGATTCTCCACACTGAGTTTAAATCT-----AAGGG 1290
 Db 3945 sValLeuGln---ThrHisCysAspCysLysAspAlaGluGlyGlyLeuValProAlaGl 3964
 QY 1291 ATTGAATAATGGAGTAGAGTATAAGTGCAATGTCATGTTGATTTGCTCTAGTAGAAA 1350
 Db 3964 yLys-ThrTrpThrSerLysAspCysThrGlnSerCys---AlaCysThrGlyGlyAlaV 3983
 QY 1351 TCTCTAGTTTAGAAAAGGATGTTTAGGGAAACATGAGGCTGCTGCGACCAACACACAGG 1410
 Db 3983 aLgln-----CysGlnAsnPheGlnC 3990
 QY 1411 CTCCCTCGATCCCTGGGCCCCAGGGAGTTTACTCAGAGCTCTCTGAAGATGGCAACCC 1470
 Db 3990 ysProLeu-----GlyThrTyrrCysLysAspSerGlyAspGlySerA 4005
 QY 1471 ATGCCCTTTTCTGAGGAGTGCATGGCTGAGCATGTTTCTGCTGCGCCAGAGAGAG 1530
 Db 4005 snCys-----ThrLysIleHis-----LysGlyA 4013
 QY 1531 AGCTTGGGTTCCTCATAGTCTGGGAGAGTGTCTGCGAGGCGCGGAGGAGCAGACA-----1586
 Db 4013 laMetGly-----AspGlyValLeuMetAlaGlyGlyIleArgAlaLeuG 4028

Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
|||
Qy 797 AAGCATGCTGTGTAGAAATCCAGAGCTTCCAGCCAAATATGGGAGGCCAATTG 856
|||
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
|||
Qy 857 GGC-----AATGTCAGAGCCCGACACACAAAGGGTCACTCGAGGAGTCCAGGC 907
|||
Db 1497 GlyIleThrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
|||
Qy 908 TGCCTGAAGAAT-----TACTGGAGTGTATGAGGCCCAAAATATGTTCTTCTATT 961
|||
Db 1511 MetPheAspSerValLeuGluCysTyrAspAla----- 1523
Qy 962 TGCAAATGCATGTGTTCCAAATATGAGAAAGCCAGACGAAAGACACATTAATGAGC 1021
|||
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro-----AlaGluThrThrValArg 1538
|||
Qy 1022 ATGCCAACTACATGCAGACTGGAGGTTG----- 1051
|||
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
|||
Qy 1052 GAAGC-----AGCATTTACCTGCCACCAAGCAATTT 1084
|||
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1578
|||
Qy 1085 TCAGGA-----CTTCCAAGATTCAGTC-----ACGATAGCGGCTCTCTCA 1126
|||
Db 1579 SerGlyGlnAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
|||
Qy 1127 TGCATCTCTCTGGAGGTG-----TGGAGGCCACATCGCCTCGCTGCTGCTCAG 1177
|||
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 1613
|||
Qy 1178 GGAGAGAGCCGGAAGAACACTGCTCCAGTGCCTGCAGACAGATGATCCTGGAG 1337
|||
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
Qy 1238 GAATTTGGAAGGTGCTTATCACGATTCCTCCACATGAGT----- 1277
|||
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
Qy 1278 -----TTAATCTAAGGATTTGAAATAGTAGTAGATATAAG 1315
|||
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 1665
|||
Qy 1316 TGTGAATGCA-----TGTTGATTTCTCTAGTCTAGAAATCTCTAGTTTAGAAAGGAT 1369
|||
Db 1665 yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
Qy 1370 GTTTAGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGCATCCCTGGGC 1429
|||
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArg 1698
Qy 1430 CCAGGAGTTTACTAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTC 1483
|||
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
|||
Qy 1484 TGAGGAGTGCATGCTGAGCTGAGCTGTTTCTGTGCGCCAGAGGAGAGCTTGGTTCCC 1543
|||
Db 1718 eGluGln-----GlyMetMetLeuAlaGluInPheLysGlnLysAlaLeuGlyLeu- 1735
|||
Qy 1544 ATAGTCTGGGAGAGTGTCTGCAGGCGGCGGAGGCGGAGGCGGCTCGGAGAGCTT 1603
|||
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
|||
Qy 1604 ACTCTGCTGAGCTCTT 1619
|||
Db 1752 rAsnTrpGlnLysLeu 1757
|||
RESULT 11
LMA5_MOUSE

ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC 061001; Q9JHQ6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpi R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bound
to each other by disulfide bonds into a cross-shaped molecule
comprising one long and three short arms with globules at each
end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ293593; CAB99255.1; -
DR EMBL; U37501; AAC53430.1; -
DR HSSP; P02468; ITLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.

```

DR Pfam: PF00055; laminin_Nterm; 1.
DR ProDom: PD002082; LamNT; 1.
DR ProDom: PD003031; Laminin_B; 1.
DR PROSITE: PS00022; EGF_1; 19.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE: PS50025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 856 1442 LAMININ DOMAIN IV 1. (DOMAIN IV B).
FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
FT DOMAIN 2169 2735 DOMAIN II AND I.
FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
FT DOMAIN 2947 3119 LAMININ G-LIKE 2.
FT DOMAIN 3120 3296 LAMININ G-LIKE 3.
FT DOMAIN 3337 3511 LAMININ G-LIKE 4.
FT DOMAIN 3512 3689 LAMININ G-LIKE 5.
FT DOMAIN 3690 2257 COILED COIL (POTENTIAL).
FT DOMAIN 2258 2464 COILED COIL (POTENTIAL).
FT DOMAIN 2465 2604 COILED COIL (POTENTIAL).
FT DOMAIN 2605 2639 COILED COIL (POTENTIAL).
FT SITE 1723 1725 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1839 1841 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 305 314 BY SIMILARITY.
FT DISULFID 307 327 BY SIMILARITY.
FT DISULFID 329 338 BY SIMILARITY.
FT DISULFID 341 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 401 410 BY SIMILARITY.
FT DISULFID 413 431 BY SIMILARITY.
FT DISULFID 434 445 BY SIMILARITY.
FT DISULFID 436 452 BY SIMILARITY.
FT DISULFID 454 463 BY SIMILARITY.
FT DISULFID 466 476 BY SIMILARITY.
FT DISULFID 500 512 BY SIMILARITY.
FT DISULFID 502 521 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 535 544 BY SIMILARITY.
FT DISULFID 547 559 BY SIMILARITY.
FT DISULFID 549 566 BY SIMILARITY.
FT DISULFID 568 577 BY SIMILARITY.
FT DISULFID 580 590 BY SIMILARITY.
FT DISULFID 593 605 BY SIMILARITY.
FT DISULFID 595 611 BY SIMILARITY.
FT DISULFID 613 622 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.

638 650 FT DISULFID BY SIMILARITY.
640 656 FT DISULFID BY SIMILARITY.
658 667 FT DISULFID BY SIMILARITY.
670 680 FT DISULFID BY SIMILARITY.
683 695 FT DISULFID BY SIMILARITY.
685 702 FT DISULFID BY SIMILARITY.
704 713 FT DISULFID BY SIMILARITY.
716 726 FT DISULFID BY SIMILARITY.
1443 1455 FT DISULFID BY SIMILARITY.
1445 1462 FT DISULFID BY SIMILARITY.
1464 1473 FT DISULFID BY SIMILARITY.
1476 1486 FT DISULFID BY SIMILARITY.
1533 1548 FT DISULFID BY SIMILARITY.
1535 1555 FT DISULFID BY SIMILARITY.
1557 1566 FT DISULFID BY SIMILARITY.
1569 1579 FT DISULFID BY SIMILARITY.
1582 1594 FT DISULFID BY SIMILARITY.
1584 1601 FT DISULFID BY SIMILARITY.
1603 1612 FT DISULFID BY SIMILARITY.
1615 1630 FT DISULFID BY SIMILARITY.
1665 1874 FT DISULFID BY SIMILARITY.
1867 1881 FT DISULFID BY SIMILARITY.
1884 1893 FT DISULFID BY SIMILARITY.
1896 1912 FT DISULFID BY SIMILARITY.
1915 1930 FT DISULFID BY SIMILARITY.
1917 1939 FT DISULFID BY SIMILARITY.
1941 1950 FT DISULFID BY SIMILARITY.
1953 1968 FT DISULFID BY SIMILARITY.
1971 1986 FT DISULFID BY SIMILARITY.
1973 1993 FT DISULFID BY SIMILARITY.
1996 2005 FT DISULFID BY SIMILARITY.
2008 2022 FT DISULFID BY SIMILARITY.
2072 2083 FT DISULFID BY SIMILARITY.
2074 2090 FT DISULFID BY SIMILARITY.
2092 2101 FT DISULFID BY SIMILARITY.
2104 2116 FT DISULFID BY SIMILARITY.
2119 2126 FT DISULFID BY SIMILARITY.
2121 2133 FT DISULFID BY SIMILARITY.
2135 2144 FT DISULFID BY SIMILARITY.
2147 2166 FT DISULFID BY SIMILARITY.
2169 2192 FT DISULFID INTERCHAIN (PROBABLE).
2172 2172 FT DISULFID INTERCHAIN (PROBABLE).
100 100 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
148 148 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
248 248 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
383 383 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
457 457 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
485 485 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
905 905 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
926 926 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
964 964 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
1335 1335 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
1534 1534 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
2021 2021 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).

Alignment Scores:
Pred. No.: 0.337 Length: 3718
Score: 117.00 Matches: 112
Percent Similarity: 29.76% Conservative: 52
Best Local Similarity: 20.33% Mismatches: 179
Query Match: 3.06% Indels: 208
DB: 1 Gaps: 31

US-09-743-237-3 (1-2134) x LMA5_MOUSE (1-3718)
QY 18 AAGCTCGCGGGGCGACGACGCGGGGAG-----CTCTCGGGGAGTAC 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleGlnGlyGlnTyr 107
QY 66 CCCGGGATC-----CCAGAGCTCAGCGGTGGAG 95
      |||
Db 108 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProValSerAsnAlaIleAsp 127
      |||

```

QY 96 GACGTC-----GCCTCTCTGAGCCCGCAGCCCGCCGCTGC 134
Db 128 GlyThrGluArgTrpTrpGlnSerProLeuSerArgGlyLeuGluTyrAsnGluVal 147
QY 135 AACGTG-----CACTCTCTGCTCGCTGCTACCGCGCAC 170
Db 148 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 167
QY 171 CGCAGCCCGCGGGTGTTCCTCCCTGGCGCTGCTGCGAAGGAGCCCTCCACCC 230
Db 168 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerThrAspPhe 183
QY 231 GGGCGTCCGCATGATCCCATGTAAGTCAAGTAAGTCAAGTCAAGTAATA 290
Db 184 GlyHisThrTyrGlnPro-----TrpGlnPhePhe 193
QY 291 ATCCGGAAGAGCAACTTTCAGNATCTTCTGCTCAGGAATCCTGTTCAGAGTTCCCAT 350
Db 194 AlaSerSerLysArgAspCys-----LeuGluArgPhe 204
QY 351 GGTCCAGGAAGTACAGGATCCCTGCTCTTCTTAAGAAAGATTCACACCCATGG 410
Db 205 GlyProArgThr-----Leu-GluArgIleThrGlnAspAs 216
QY 411 TGAT-----ATCCCAATTGAAGGGGCACACAAA 440
Db 216 pAspValIleCysThrThrGluTyrSerArgIleValProLeuGluAsnGlyGluIle-V 236
QY 441 TGCTATCATAGACAATCTAGACAGAGACTAAAGCACTCCATTTGGTTCCTCAGT 500
Db 236 alValSerLeuValAsnGlyArgProGlyAlaLeu----- 247
QY 501 ATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTTAAACCAATGACTGCTTTAGTAG 560
Db 248 -----AsnPheSerTyrSerProLeuLeuArgAspPheThrLysAlaThrA 263
QY 561 GG-----AGATTTTCCGAGCATCAACAATA-----AATCTCATACACAACAAC 608
Db 263 snIleArgLeuArgPheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaL 283
QY 609 TTGAGGAGCCTTACCATCGGTAGTCAACGGTCTGCTTCCCTCGGGATCAACTCTTC 668
Db 283 euArgAsp-----ProThrValThrArgArgTyrTyr----- 294
QY 669 CAGGACCACCAAAATAACTTTGGCTGGTACTGCTGCTTTCGCGAGTGGGACTTTT 728
Db 295 --SerIleLysAspIleSerIleGlyArgCysValCysHisGlyHisAlaAspValC 314
QY 729 GC-----AACAACT 737
Db 314 ysAspAlaLysAspProLeuAspProPheArgLeuGlnCysAlaCysGlnHisAsnThrC 334
QY 738 GC-----AATTGTAATAATTGTGCAACACTTGCATCATGATATATGAACGGTTAAAG 791
Db 334 ySGlyGlySerCysAspArgCysCysProGlyPheAsnGln-----GlnProTrpLysP 352
QY 792 CCATTAAGGCA-----TGTCTTGGT----- 811
Db 352 roAlaThrThrAspSerAlaAsnGluCysGlnSerCysAsnCysHisGlyHisAlaTyrA 372
QY 812 -----AGAAATCCAGAGCTTTCACGCCCAAAATTTG 842
Db 372 spCysTyrTyrAspProGluValAspArgArgAsnAlaSerGlnAsnGlnAspAsnValT 392
QY 843 GGAAGGCCCAATTGGCAATGTC-----AAGCCCCAGCACACAAAGGGTGCACAT 893
Db 392 yrGlnGly-----GlyGlyValCysLeuAspCysGlnHisHisThrGlyIleAsnC 410
QY 894 CGAGGAGGTGAGGTGCTGAAGAATTAATCTGCGAGTGCTATGAGGCCCAATATGTGTT 953
Db 410 ySGluArg-----CysLeuProGlyPhePheArgAlaProAspGlnProLeuAspSerP 428
QY 954 CTTCATTATTCAAATGCAATTCATTTGGTTCACAAAT-----TATGAAGAAA 995

Db 428 roHisValCysArgProCysAspCysGluSerAspPheThrAspGlyThrCysGluAspL 448
QY 996 GCCCAGAACCAAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGGTTTG--- 1051
Db 448 euThrGlyArg-----CysTyrCysArgProAsnPhe-----ThrGlyGluLeuCysA 464
QY 1052 -----GAAGGC-----AGCCATTACCTGCCACCAACAAATTTTCAG 1088
Db 464 laAlaCysAlaGluGlyTyrThrAspPheProHisCysTyrPro----- 478
QY 1089 GACTTCCAGATTTCAGTCAC---GATAGCGCGCTTCTCTCATCATCTCTCTGGAGGTGG 1145
Db 479 --LeuProSerPheProHisAsnAspThrArgGluGlnValLeuProAlaGlyGlnIleV 498
QY 1146 TGGAGGCCACATCGCTGCTGCTCAGGGAAGAGCGCGGAGAAAGAA--- 1198
Db 498 alAsnCysAspCysAsnAlaAlaGlyThrGlnGlyAsnAlaCysArgLysAspProArgL 518
QY 1199 -----CACTGCTCCAACTGC----- 1213
Db 518 euGlyArgCysValCysLysProAsnPheArgGlyAlaHisCysGluLeuCysAlaProG 538
QY 1214 -----CTGGCAGAGCAGA 1226
Db 538 lyPheHisGlyProSerCysHisProCysGlnCysSerSerProGlyValAlaAsnSerL 558
QY 1227 TGATCTCTGGAGGAATTTGGAAGGTGCTTA 1255
Db 558 euCysAspProGluSerGlyGlnCysMet 567
RESULT 12
NTC4_HUMAN STANDARD: PRT: 2003 AA.
ID NTC4_HUMAN Q99466; Q00306; Q99490; Q99458; Q9H3S8; Q9U1I9; Q9U1J0;
AC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE (hNotch4).
GN NOTCH4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
RC TISSUE-Placenta;
RX MEDLINE=973111416; PubMed=9168133;
RA Sugaya K., Sasunuma S.-I., Nohata J., Kimura T., Fukagawa T.,
RA Nakamura Y., Ando A., Inoko H., Ikenura T., Mita K.;
RT "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
RT human counterpart gene of mouse proto-oncogene Int3";
RL Gene 189:235-244(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE-Bone marrow, and Heart;
RX MEDLINE=98360091; PubMed=9693032;
RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
RT "Cloning, characterization, and the complete 56.8-kilobase DNA
RT sequence of the human NOTCH4 gene";
RL Genomics 51:45-58(1998).
RN [3]
RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
RA Miyagawa T., Tokunaga K., Hojho H.;
RT "Human notch4 gene variant";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L., S.;
RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -|- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -|- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -|- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
 CC produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -|- PM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -|- PM: Phosphorylated (By similarity).
 CC -|- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -|- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -|- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -|- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -|- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D63395; BAA09708.1; ALT_FRAME.
 CC EMBL; D86566; BAA13116.1; -.
 CC EMBL; U95299; AAC3288.1; -.
 CC EMBL; U89335; AAC63097.1; -.
 CC EMBL; AB023961; BAB20317.1; -.
 CC EMBL; AB024520; BAA88951.1; -.
 CC EMBL; AB024578; BAA88952.1; -.
 CC HSP; P08709; I8F9.
 CC Genew; HGNC:7884; NOTCH4.
 CC MIM; 164951; -.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR000742; EGF-2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF-II.
 CC InterPro; IPR000800; Notch.
 CC Pfam; PF00008; EGF; 26.
 CC Pfam; PF00023; EGF; 6.
 CC Pfam; PF00066; notch; 2.
 CC PRINTS; PR00010; EGFBLD.

DR PRINTS; PR00011; EGFAMININ.
 DR PRINTS; PR00012; FNTYPEI.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00001; EGF_Like; 15.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 2003 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
 FT CHAIN 1432 2003 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT SIMILARITY).
 FT CHAIN 1467 2003 NOTCH INTRACELLULAR DOMAIN (BY
 FT SIMILARITY).
 FT DOMAIN 24 1447 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1448 1468 POTENTIAL.
 FT DOMAIN 1469 2003 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 24 63 EGF-LIKE 1.
 FT DOMAIN 64 115 EGF-LIKE 2.
 FT DOMAIN 118 135 EGF-LIKE 3.
 FT DOMAIN 156 192 EGF-LIKE 4.
 FT DOMAIN 194 232 EGF-LIKE 5.
 FT DOMAIN 234 274 EGF-LIKE 6.
 FT DOMAIN 276 312 EGF-LIKE 7.
 FT DOMAIN 314 353 EGF-LIKE 8.
 FT DOMAIN 355 391 EGF-LIKE 9.
 FT DOMAIN 392 430 EGF-LIKE 10.
 FT DOMAIN 432 473 EGF-LIKE 11.
 FT DOMAIN 475 511 EGF-LIKE 12.
 FT DOMAIN 513 549 EGF-LIKE 13.
 FT DOMAIN 551 587 EGF-LIKE 14.
 FT DOMAIN 589 625 EGF-LIKE 15.
 FT DOMAIN 626 659 EGF-LIKE 16.
 FT DOMAIN 661 689 EGF-LIKE 17.
 FT DOMAIN 691 727 EGF-LIKE 18.
 FT DOMAIN 729 765 EGF-LIKE 19.
 FT DOMAIN 767 803 EGF-LIKE 20.
 FT DOMAIN 806 842 EGF-LIKE 21.
 FT DOMAIN 844 880 EGF-LIKE 22.
 FT DOMAIN 882 928 EGF-LIKE 23.
 FT DOMAIN 930 966 EGF-LIKE 24.
 FT DOMAIN 968 1004 EGF-LIKE 25.
 FT DOMAIN 1006 1044 EGF-LIKE 26.
 FT DOMAIN 1046 1085 EGF-LIKE 27.
 FT DOMAIN 1087 1126 EGF-LIKE 28.
 FT DOMAIN 1130 1171 EGF-LIKE 29.
 FT DOMAIN 1172 1212 POLY-ARG.
 FT REPEAT 1165 1212 LIN/NOTCH 1.
 FT REPEAT 1213 1246 LIN/NOTCH 2.
 FT REPEAT 1247 1286 LIN/NOTCH 3.
 FT REPEAT 1633 1665 ANK 1.
 FT REPEAT 1666 1698 ANK 2.
 FT REPEAT 1700 1732 ANK 3.
 FT REPEAT 1733 1765 ANK 4.
 FT REPEAT 1766 1798 ANK 5.
 FT DISULFID 28 41 BY SIMILARITY.
 FT DISULFID 35 51 BY SIMILARITY.
 FT DISULFID 53 62 BY SIMILARITY.
 FT DISULFID 68 80 BY SIMILARITY.
 FT DISULFID 74 103 BY SIMILARITY.
 FT DISULFID 105 114 BY SIMILARITY.
 FT DISULFID 122 133 BY SIMILARITY.
 FT DISULFID 127 143 BY SIMILARITY.
 FT DISULFID 145 154 BY SIMILARITY.
 FT DISULFID 160 171 BY SIMILARITY.

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11105;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RN SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomsky Z.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58335; AA472945.1; -
DR PIR: A38465; GNWVTC.
DR PDB: 1AIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; -
DR MEROPS: U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF00998; HCV_RDRP; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1852 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; P8422D5ECCFDF9C CRC64;

Alignment Scores:

Pred. No.:	0.386	Length:	3010
Score:	116.00	Matches:	93
Percent Similarity:	34.88%	Conservative:	57
Best Local Similarity:	21.63%	Mismatches:	157
Query Match:	3.03%	Indels:	123
DB:	1	Gaps:	23

US-09-743-237-3 (1-2134) x POLG_HCVBK (1-3010)

QY	542	CCAATGACGCTTTAGTAGGAGA-----	TTTTTGCAGCATCAACAAATTA	589
Db	1381	ProileglualaleargGlyGlyArgHisLeullepheCysHisSerLysLysCys		1400
QY	590	AATCTCATTACACAACAACCTTGAGGCGCTTACCATCGGTGACACGGGTCTGCTTTC		649

Db 1401 AspGluLeuAlaAlaLysLeuSerGlyLeu-----GlyIleAsnAlaValAlaTyr 1417
QY 650 CCCTCGGGATCAACTCTTCAGGACCAACCAAAATA----- 685
Db 1418 TyrArgGlyLeuAspValSerValIleProThrIleGlyAspValValValAlaThr 1437
QY 686 -----ACTTTGGCTGGGTACTGTGACTGCTTGCAGTGGGACTTTTGCACAACTGC 739
Db 1438 AspAlaLeuMetThrGlyTyr-----ThrGlyAspPheAspSerValIle 1452
QY 740 AATGTAAATAATGTTCCCACAACTTG-----CATCATGATATT 778
Db 1453 AspCysAsnThrCysValThrGlnThrValAlaPheSerLeuAspProThrPheThrIle 1472
QY 779 GAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGAAATCCAGACCAAAA 838
Db 1473 GluThrThrThrValProGlnAspAlaValSerArgSerGlnArg-----ArgGlyArg 1490
QY 839 ATTGGGAAGGCCAATTGGGC-----AATGTCAGGCCCCAGCAGCAACAAGGGTGC 889
Db 1491 ThrGlyArgGlyArgArgGlyIleTyrArgPheValThrProGlyGlu----- 1506
QY 890 AACTGCAGGAGGCTGAGCTCCCTCAAGAAAT-----TACTGCCAGTCTATGAGGCCAA 943
Db 1507 -----ArgProSerGlyMetPheAspSerValLeuCysGluCysTyrAspAla--- 1523
QY 944 ATTATGTGTTCTTCTATTGCAATGATGTTGTTGCAAAAATATGAAGAAAGCCAGAA 1003
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro--- 1532
QY 1004 CGAAGACACTAATGACATGCCAACTACATGCAGCTGAGGTTTGGAA----- 1054
Db 1533 AlaGluThrSerValArgLeuAlaGlyAlaTyrLeuAsnThrProGlyLeuProValCysGln 1552
QY 1055 -----GCGAGCCATTAC 1066
Db 1553 AspHisLeuGluPheTrpGluSerValPheThrGlyLeuThrHisIleAspAlaHisPhe 1572
QY 1067 CTGCCACCAACGAAATTTTCAGGA-----CTTCCAGATTCAGTC-----AC 1108
Db 1573 LeuSerGlnThrLysGlnAlaGlyAspAsnPheProThrLeuValAlaTyrGlnAlaThr 1592
QY 1109 GATAGGCGGCTTCTCATGATCTCTCTGGAGGTGG-----TGGAGGCCACATGC 1159
Db 1592 rValCysAlaArgAlaGlnAlaProProSerTrpAspGlnMetTrpLys----- 1609
QY 1160 GCCTGCTGCTGTCTCAGGAGAGAGCGCGAGAAAGAACTGCTCCAAGTGCCTGCGCA 1219
Db 1610 -----CysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrAr 1627
QY 1220 GAGCAGATGATCCTGGAGGAATTTGGAAGGTGCTTATCACAGATCTCCACACTGAGT-- 1277
Db 1627 gLeu-----GlyAlaValGlnAsnGluValThrLeuThrHi 1639
QY 1278 -----TTAAATCAAGGATTCGAAA 1297
Db 1639 sProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrSerTh 1659
QY 1298 ATGGAGTAGAGTAAAGTCTGAATGATGCTGATTTTGTCTAGTACAAATCTCTAG 1357
Db 1659 rTpValLeuValGlyValLeuAlaLeuAlaAlaTyrCysLeuThrThrGlySe 1679
QY 1358 TTTAGAAGGATGTTTGGGGAAACATGAGGCTGGCTTCGACGCAACACAGGCTCCCTC 1417
Db 1679 rValValIleVal-----GlyArgIleIleLeuSerGlyArgProAlaIle 1694
QY 1418 GCATCCCTGGCCCGGAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGCAACCCA 1471
Db 1694 eValProAspArgGluLeuLeuTyrGlnGluPheAspGluMetGluGluCysAlaSerHi 1714
QY 1472 TGCCCCCTTTTCTGAGGAGGTGCATGGCTGCCTGAGCATGTTTGTCTGCCCGAGGAGAGA 1531
Db 1714 sLeuProTyrIleGluGln-----GlyMetGlnLeuAlaGluGlnPheLysGlnLysAl 1732

QY 1532 GCTTGGGTCCCATAGTCTCTGGGAGAGTGTCTGCAGGGCGCGGAGGCGAGAGAGGCC 1591
Db 1732 aLeuGlyLeu-----LeuGlnThrAlaThrLysGlnAlaGluAla---AlaAlaPr 1748
QY 1592 TCGCGAGAGCTCACTCTGGTCGACTCTT 1619
Db 1748 oValValGluSerLysTrpArgAlaLeu 1757
RESULT 14
POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals".
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1',
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D11168; BAA01943.1;
DR PIR: A45573; A45573.
DR HSSP: P26663; 1JXP.
DR MEROPS: S29.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Y00057; CAA68266.1; -
CC EMBL; X64072; CAA45427.1; JOINED.
CC EMBL; X64073; CAA45427.1; JOINED.
CC EMBL; X64074; CAA45427.1; JOINED.
CC EMBL; X64075; CAA45427.1; JOINED.
CC EMBL; X64076; CAA45427.1; JOINED.
CC EMBL; X64077; CAA45427.1; JOINED.
CC EMBL; X64078; CAA45427.1; JOINED.
CC EMBL; X64079; CAA45427.1; JOINED.
CC EMBL; X64080; CAA45427.1; JOINED.
CC EMBL; X64081; CAA45427.1; JOINED.
CC EMBL; X64082; CAA45427.1; JOINED.
CC EMBL; X64083; CAA45427.1; JOINED.
CC EMBL; X63924; CAA45427.1; JOINED.
CC EMBL; X63925; CAA45427.1; JOINED.
CC EMBL; X63926; CAA45427.1; JOINED.
CC EMBL; AL163300; CAB90553.1; -
CC EMBL; M15395; AAS59490.1; -
CC EMBL; S81234; AAB21404.1; -
CC PIR; A25967; IJHULM.
CC PIR; A29265; A29265.
CC HSSP; P05106; IJV2.
CC Genew; HGNC:6155; ITGB2.
CC MIM; 600065; -
CC MIM; 116920; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002369; Integrin_B.
CC InterPro; IPR001169; Integrin_beta_C.
CC InterPro; IPR003659; Plexin-like.
CC InterPro; IPR002035; WVF-A.
CC Pfam; PF00362; Integrin_B; 1.
CC PRINTS; P01186; INTEGRINB.
CC ProDom; PD001811; Integrin_B; 1.
CC SMART; SM00187; INB; 1.
CC SMART; SM00423; PST; 1.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00243; INTEGRIN_BETA; 3.
CC PROSITE; PS00022; EGF_1; UNKNOWN_2.
CC PROSITE; PS01186; EGF_2; UNKNOWN_3.
CC Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal; Disease mutation.
FT SIGNAL 1 22
FT CHAIN 23 769 INTEGRIN_BETA-2.
FT DOMAIN 23 700 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 701 723 POTENTIAL.
FT DOMAIN 724 769 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 124 363 WVF-A-LIKE.
FT DOMAIN 449 617 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 449 496 I.
FT REPEAT 497 540 II.
FT REPEAT 541 581 III.
FT REPEAT 582 617 IV.
FT DISULFID 25 447 BY SIMILARITY.
FT DISULFID 33 43 BY SIMILARITY.
FT DISULFID 36 73 BY SIMILARITY.
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 191 198 BY SIMILARITY.

Alignment Scores:
Pred. No.: 0.32 Length: 769
Score: 115.50 Matches: 59
Percent Similarity: 34.16% Conservative: 37
Best Local Similarity: 21.00% Mismatches: 101
Query Match: 3.02% Indels: 84
DB: 1 Gaps: 16

US-09-743-237-3 (1-2134) x ITB2_HUMAN (1-769)

QY 304 AACTTTCAGAAATCTTCTGCTCAGGAATCCTGTGCAAGTTCCTCATGTCCTCCAGGAACT 363
Db AsnGlnProArgGlyAspCysAspGlyValGlnIleAsnValProIle-ThrPheGlnVa 413
QY 364 AGAGGATCCCTCTGCTGTTCTCTTAAGAAAGATTCC----- 400
Db 1LysValThrAlaThrGluCysIleGlnGluInSerPheValIleArgAlaLeuGlyPh 433
QY 401 -AACCCCAATGGTGATATGCAATTTGAAAGGGGGCACAAATGCTATGTATGATAGCAATTC 459
Db 433 eThrAspIleValThrValGlnValLeuProGlnCysGluCysArgCysArgaspGlnSe 453
QY 460 TAGACAAGAGAACTAAAGCACTCCATTGGTTCTCAGTATCAAGATCAAAATAATTA 519
Db 453 rArgAspArgSerLeu-----CysHis-----GlyLysGlyPh 464
QY 520 TCTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGAGAGATTTTGGCAGCATC 579
Db 464 eLeuGluCysGlyIleCysArgCysAspThrGlyTyrIleGlyLysAsnCysGluCysG1 484
QY 580 AACAAAATTAATCTCATTACACAACACTTGAGGGAGCC----- 619
Db 484 nThrGln---GlyArgSerSerGlnGluLeuGluGlySerCysArgLysAspAsnAsnSe 503
QY 620 -----TTACCATCGGTAGTCAACGGGCTCTGCTTCCCTCGGATCAAC 663
Db 503 rIleIleCysSerGlyLeuGlyAspCysValCysGlyGlnCysLeuCysHisThrSerAs 523
QY 664 TCTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGACTGTGCTTTGCC----- 715
Db 523 pValProGly-----LysLeuIleTyrGlyGlnTyrCysGluCysAspThrIleAsnCy 541
QY 716 -----AGTGGGGACTTTTCCCAACAC-----TGCAATTGTAATAA 750
Db 541 sGluArgTyrAsnGlyGlnValCysGlyGlyProGlyArgGlyLeuCysPheCysGlyLy 561
QY 751 TTGTTGCACAACTTGTCATCATGATATTGAACGGTTT-----AAAGCCAT 795
Db 561 sCys-----ArgCysHisProGlyPheGluGlySerAlaCysGlnCysGluArgThrTh 579
QY 796 TAAGGATGCTCTGGTAGAAATCCAGAGCTTTCCAGCCCAAAATTTGGGAGGGCCCAATT 855
Db 579 rGluGlyCysLeu-----AsnProArgArgValGluCysSer---GlyArgGlyArg-- 595
QY 856 GGCAATGTCAAGCCCCAGCACAAAGGGTGCACACTGCAGGAGGTGCAGCTCCCTGAA 915
Db 596 -----CysArgCys----- 598
QY 916 GAATTACTGCGAGTGTATGAGGCCCAAAATTATG-----TG 951
Db 599 -AsnValCysGluCysHisSerGlyTyrGlnLeuProLeuCysGlnGluCysProGlyCy 618
QY 952 TTCTTCTATTTC-----AAATGCATTGGTTGCAAAATTTATGAAGAAAGCC 999
Db 618 sProSerProCysGlyLysTyrIleSerCysAlaGluCysLeuLysPheGluLysGlyPr 638
QY 1000 A 1000
Db 638 o 638

Search completed: April 21, 2003, 11:57:01
Job time : 90.5 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:50:49 ; Search time 136.5 Seconds
(without alignments)
6442.557 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcgaaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USTPO.spool/US0974337/runat_21042003_112209_1473/app-query.fasta_1.2311
-DB=SPTREMBL_21 -Qfmt=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0974337 -CGN_1_1.238 -runat_21042003_112209_1473 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databse :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515.5	39.6	475	11 Q8VIE0	Q8vie0 mus musculus

2	1508.5	39.4	475	11 Q8VIE1	Q8vie1 mus musculus
3	633.5	16.6	277	11 Q9D571	Q9d571 mus musculus
4	483.5	12.6	950	5 Q9V608	Q9v608 drosophila
5	444.5	11.6	435	5 Q95QD7	Q95qd7 caenorhabdi
6	437	11.4	429	5 Q62295	Q62295 caenorhabdi
7	376	9.8	571	10 Q9SL70	Q9sl70 arabidopsis
8	358	9.4	603	10 Q9SAD1	Q9sca1 arabidopsis
9	264.5	6.9	601	10 Q9CAV1	Q9cav1 arabidopsis
10	262.5	6.9	356	10 Q9LW71	Q9lw71 arabidopsis
11	248.5	6.5	896	10 Q9ZS22	Q9zs22 glycine max
12	248	6.5	658	10 Q23333	Q23333 arabidopsis
13	247.5	6.5	609	10 Q9M679	Q9m679 arabidopsis
14	243.5	6.4	695	10 Q9LE32	Q9le32 arabidopsis
15	243.5	6.4	695	10 Q9LUI3	Q9lui3 arabidopsis
16	242.5	6.3	526	10 Q9A12	Q9a12 arabidopsis
17	202.5	5.3	593	10 Q9LUI5	Q9lui5 arabidopsis
18	200	5.2	243	5 Q9VMQ3	Q9vmq3 drosophila
19	192	5.0	553	10 Q94DS2	Q94ds2 oryza sativ
20	147	3.8	1704	5 Q94446	Q94446 chironomus
21	137	3.6	4123	4 Q75851	Q75851 homo sapien
22	136.5	3.6	1637	6 Q9XSV8	Q9xsv8 bos taurus
23	136.5	3.6	5146	6 Q8SPM4	Q8spm4 bos taurus
24	132.5	3.5	1237	11 Q9D2D7	Q9d2d7 mus musculus
25	130.5	3.4	3695	4 Q8TDF8	Q8tdf8 homo sapien
26	130	3.4	2327	13 Q91BG7	Q91bg7 xenopus lae
27	127.5	3.3	3010	12 Q9DTE1	Q9dte1 hepatitis c
28	127	3.3	3010	12 Q9DTE7	Q9dte7 hepatitis c
29	125.5	3.3	5374	11 Q99ND0	Q99nd0 mus musculus
30	125	3.3	1028	11 Q9JLL0	Q9jll0 mus musculus
31	124.5	3.3	1172	10 Q8RVG7	Q8rvg7 arabidopsis
32	124	3.2	969	4 Q96KG6	Q96kg6 homo sapien
33	122.5	3.2	3198	5 Q9U8G8	Q9u8g8 manduca sex
34	122	3.2	3010	12 Q9Q1Y8	Q9q1y8 hepatitis c
35	122	3.2	3010	12 Q93JG6	Q93jg6 hepatitis c
36	121.5	3.2	3010	12 Q9QIX2	Q9qix2 hepatitis c
37	121	3.2	1737	4 Q75097	Q75097 homo sapien
38	120.5	3.2	1587	4 Q00508	Q00508 homo sapien
39	120	3.1	3010	12 Q81825	Q81825 hepatitis c
40	120	3.1	3010	12 Q68285	Q68285 hepatitis c
41	120	3.1	3010	12 Q99AU2	Q99au2 hepatitis c
42	120	3.1	3011	12 Q9DTE3	Q9dte3 hepatitis c
43	120	3.1	3014	12 Q9DTE0	Q9dte0 hepatitis c
44	119.5	3.1	1035	5 Q9VJU8	Q9vjv8 drosophila
45	119.5	3.1	1511	4 Q75412	Q75412 homo sapien

ALIGNMENTS

RESULT 1

ID	Q8VIE0	PRELIMINARY;	PRT;	475 AA.
AC	Q8VIE0			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Tesmin (Hypothetical 50.6 kda protein).			
GN	TESM.			

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=BALE/C;

RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
protein; a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[2]

RN SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.


```

Db 162 erSerGlnGluAlaGluGluAlaSerCysProArgLysLysAspSerProMetv 182
QY 411 TGATATGCCAATTAAGAGGGGGCACCAACATGCTATGATAGACAATTTAGAACAGAG 470
Db 182 allleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArg 202
QY 471 AACTAAAGCAGCTCCATTTGGTCTCCAGTATCAAGATCAAAATATATCTACAGTCAG 530
Db 202 luLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGlnSerG 222
QY 531 ATGTCCTTAACCAATGACCTTTAGTAGGAGATTTTGGCCAGCATCAACAAATTA 590
Db 222 luLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLysLeu 242
QY 591 ATCTCATTTACACCAACTTGAGGGAGCCCTACCATCGGTAGTCAACGGGTCTGCTTTC 650
Db 242 snLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPhe 262
QY 651 CCTCGGATCACTCTCCAGGACCACCAAAATACTTTGGCTGGG 697
Db 262 roSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGly 277

RESULT 4
Q9V6Q8
ID Q9V6Q8 PRELIMINARY; PRT; 950 AA.
AC Q9V6Q8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarik C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003818; AAF58365.1; -.
DR FlyBase; FBgn0033846; CG6061.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 950 AA; 100021 MW; 627C3EA6B4A0A30 CRC64;

Alignment Scores:
Pred. No.: 1,01e-38 Length: 950
Score: 483.50 Matches: 155
Percent Similarity: 39.73% Conservative: 81
Best Local Similarity: 26.09% Mismatches: 168
Query Match: 12.64% Indels: 190
DB: 5 Gaps: 18

US-09-743-237-3 (1-2134) x Q9V6Q8 (1-950)
QY 1 AATTCGGGGTCAAGCGCAAGCTCGCGGGGGGCGACAGCGCGGGG-----ACG 51
Db 357 AsnLysAsnIleLeuLysLeuSerProAsnAlaSerAlaThrSerSerThrHisThr 376
QY 52 TCCTCGGGGAGTACCCCGGGATCCAGAGC----- 81
Db 377 ThrGlyGlnThrProSerSerGlyLeuHisAlaIleGlnLeuProGlyLys 396
QY 82 -----TCAGCGCGCTGGAGGAGCTGC 102
Db 397 GlyGlyIleGlnTyrValArgValLeuAsnAsnAsnLysSerAlaAlaGlyThrSerAla 416
QY 103 CCTCTCTCAGCGCCCGCGAGC-----CGCCCGCTGCA 135
Db 417 ThrAlaSerIleProLysThrValGlnThrGlnLysIleThrValValArgProAla 436
QY 136 -----ACGTGCACCTTCCTGCTCGCTGCTACCGCGGCACC 171
Db 437 AlaThrGlyValProAlaThrSerThrThrThrSerAlaAlaAlaSerProAlaAla 456
QY 172 GCAGCCCGGGGTGTTCCTCCCTGG----- 198
Db 457 AlaSerLysAlaAsnLeuAlaMetGlyAsnThrAsnLysIleValMetArgSerMetGly 476
QY 199 -----CGCGCTGGGTCC 210
Db 477 GlySerIleValProLeuProSerValGlnThrLeuValSerLysArgAlaLeuGlyAla 496
QY 211 TCCGAAGGAGCCTCCACCCCGCGCTCCGCATGATCC----- 247
Db 497 IleSerAsnAlaSerLysPro-AlaSerAlaAlaSerSerSerSerAlaThrProSerAlaSe 516
QY 248 -----CAGTTGAATTCAGGTA 264
Db 516 rGlnGluLeuProArgLysHisArgLeuThrAspLeuAsnValGlnLeuLysGlnSerAl 536
QY 265 AGCAGGTGTACTACTACAGTAATAAT-----CCGGAAGAAGCAACTTTGCA 312
Db 536 aserValSerSerGluAlaSerAspSerSerAspAlaGlyProGluAlaLysLysProAr 556
QY 313 GAATCTTCTGCTCAGGAATCTCTGTGCAAGTTCCTCCAGCAACTAGAG----- 367
Db 556 gTyrValIleThrMetGlnGlnGlySerGlnLysAlaAlaSerGlnProValGlnLysLe 576
QY 368 -----GATGCTCTCTGCTG 381
Db 576 uIleAsnArgThrAlaAsnValGlnArgValValSerSerSerSerThrSerProSerSer 596
QY 382 TTCTCTTAAGAAAGATTCCCAACCCCAATGTGATATGTCCTCAATTAAGAAAGGGCGCACAAAT 441
Db 596 nSerThrLysLysIleTyrAsnTyrValGlnProThrGlySerAsnGlyAlaLysTyrMe 616

```



```

QY 667 TCAGAGACCAACCAAAATTAACCTTGGCTGGTACTGCTGCTGGTCCAGTGGGACATT 726
Db 181 sserGlnCysLeuLysLeu-----TyrCysAspCysPheAlaAsnGlyLuhp 197
QY 727 TTCGCAACACTGCAATTAATATGTTGCAACAACTTGCATCATGATATCAACGGTT 786
Db 197 ecysArgAspCysAsnCysLysAspCysHisAsnAlaGluTyrAspSerGlnArgSe 217
QY 787 TAAAGCCATTAAAGCCATGCTTGGTGAATAATCCAGAGCTTTCAGCCCAAAATTTGGG-- 844
Db 217 rLysAlaIleArgGlnSerLeuGluArgAsnProAsnAlaPheLysProLysIleGlyl 237
QY 845 ----AAGGGCAATTGGCAATGTCGAAGCCAGCAGCAACAAGGTCGAAGTCAGAG 900
Db 237 eAlaArgGlyIleThrAspLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSe 297
QY 901 GTCAGCTGCTCGAACAATTAATGAGTGCAGTGCATGAGCCCAAAATTAATGCTCTCTAT 960
Db 257 sserGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysValProCysThrAsp 277
QY 961 TTGCAAAATCATGTTGCAAAAT-----TATGAAGAAAG 996
Db 277 gCysLysCysLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSe 297
QY 997 C-----CCAGAGAAAGACACTAATGAGCATGCCAACTACATGACGACT-- 1042
Db 297 rGlyGlyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAsnAlaSerSerThrAl 317
QY 1043 -----GGAGGTTG-----GAAGGCAGCATTAACCTGCC 1071
Db 317 aThrProAspSerGlyProGlySerValValThrAspGluHisGlyAspTyr---Gl 336
QY 1072 ACCAAGAAATTTTCAGGACTTCCAGATTCACAGTACAGTAGCGG-----CCTTCCTC 1125
Db 336 uAspMetLeuLeuSerHisLysProLysValGluMetAspProArgArgPheProTyr 356
QY 1126 ATGCATCTCTGGAGGTGGTGGAGCCACATCGCGCTGCTGCTGCTGCTGAGGAGAA 1185
Db 356 rTyrMetThrAspGluValGluAlaAlaThrMetCysMetValAlaGlnAlaGlu 376
QY 1186 GGCC-----GAGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGAT 1227
Db 376 uAlaLeuAsnTyrGluLysValGlnThrGluAspGluLysLeuIleAsnMetGluLysLe 396
QY 1228 GATCTCGGAGGAATTTGGAAGGTGCTTATCAGATTC 1267
Db 396 uValLeuArgGluPheGlyArgCysLeuGluGlnMetIle 409
RESULT 6
O62295 ID O62295 PRELIMINARY; PRT; 429 AA.
AC O62295;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Jc8.6b protein.
GN Jc8.6b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 282274; CAB05228.1; -.

```

```

DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

```

```

Alignment Scores:
Pred. No.: 3, 32e-34 Length: 429
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35
Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 5 Gaps: 8

```

US-09-743-237-3 (1-2134) x O62295 (1-429)

```

QY 698 TACTGTGACTGCTTCCAGTGGGACTTTTGCACAACTGCAATTTGTAATATTGCTTGC 757
Db 182 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 201
QY 758 AACAACTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATGCTTGGTAGAAAT 817
Db 202 AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 221
QY 818 CCAGAGCTTTCAGCCCAAAATTTGGG-----AAGGGCAATTGGGCAATGTCAGCCC 871
Db 222 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArg 241
QY 872 CAGCAACAACAAAGGTGCAACTGCAAGGAGTGCAGGCTGCTGCAAGAACTTACTGCGAGTGC 931
Db 242 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 261
QY 932 TATGAGGCCAAATTAATGTTGCTTCTTATTTGCAATGCAATGCTGTTGCAAAAT----- 985
Db 262 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 281
QY 986 -----TATGAAGAAAGC-----CCAGAACGAAAGACACTA 1015
Db 282 ThrTyrArgMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 301
QY 1016 ATGACATGCCAACTACATGCAGACT-----GGAGGTTG--- 1051
Db 302 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 321
QY 1052 -----GAGGCGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAAGATTC 1102
Db 322 ThrAspGluHisGlyAspTyr---GluAspMetLeuLeuSerHisLysProLysVal 340
QY 1103 AGTCACGATAGCGG-----CCTTCCTCATCATCTCTCTGGAGGTGCTGAGGCCACA 1156
Db 341 GluMetAspProArgArgPheProTyrTyrMetThrAspGluValValGluAlaAla 360
QY 1157 TCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
Db 361 ThrMetCysMetValAlaGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGlu 380
QY 1199 CACTGCTCAAGTGCCTGCGCAGCAGCAGATGATCTCTGGAGGAATTTGGAAGGTGCTTATCA 1258
Db 381 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 400
QY 1259 CAGATTCTC 1267
Db 401 GlnMetIle 403
RESULT 7
O9SL70 ID O9SL70 PRELIMINARY; PRT; 571 AA.
AC O9SL70;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AT2G20110 protein.
GN AT2G20110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Bowan C.L., Barnstead M.E., Feldblum T.V.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vankken S.E., Umayam L.J., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006081; AAD24386.1; -
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2.
 SQ SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;

Alignment Scores:
 Pred. No.: 4,06e-28 Length: 571
 Score: 376.00 Matches: 95
 Percent Similarity: 44.27% Conservative: 48
 Best Local Similarity: 29.41% Mismatches: 90
 Query Match: 9.83% Indels: 90
 DB: 10 Gaps: 10

US-09-743-237-3 (1-2134) x Q9SL70 (1-571)

QY 527 TCAGATGCTCCCTAAACCAATGACTCTTTAGTAGGAGATTTTCCAGCATCAACAAA 586
 Db 51 AlaserileProSerProileValThrValThrArgProileileThrSerGlnAlaPro 70
 QY 587 TTAATCTCATTAACA-----CAACAACCTTGAGGAGCCTTA----- 622
 Db 71 ProThrValAlaThrProileProProProGlnSerGlnGlyIleileuHisVal 90
 QY 623 -----CCATCGGTACTCAACGGGTCTGCTTCCCTCCGCGATCACT 664
 Db 91 ProfileArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 110
 QY 665 CTTCCAGCACCACCAAAA-----ATAACTTTGGCT 694
 Db 111 ArgaspGlyThrProGlnLysLysLysLysLysLysLysHisSerArgCysLeuLys 130
 QY 695 GGGTACTGCTGCTGCTTCCAGTGGGACTTTTGCACAACTGCAATTTGAATAATTGT 754
 Db 131 LeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCysAsnLysLys 150
 QY 755 TGCACAACTTGCATCATGATTTGAACGGTTTAAAGCCATTAGGCATGCTTGGTAGA 814
 Db 151 PheAsnAsnValGluAsnGluProAlaArgArgGlnAlaValGluSerThrLeuGluArg 170
 QY 815 AATCCAGAAGCTTTCCACGCAAAATT-----GGGAAG 847
 Db 171 AsnProasnAlaPheArgProLysIleAlaAlaSerProHisGlyArgAspAsnArg 190
 QY 848 GCCCAATTGGCAATGTC-----AAGCCCGACCAACAAGGTGCAACTGCAGGAGG 901
 Db 191 GluGluValGlyAspValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 210
 QY 902 TCAGGCTCCCTGAACAATTACTGGAGTGTATGAGGCCCAAAATTATGTTCTTCTATT 961
 Db 211 SerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 230

QY 962 TGCRAATGCATTGTTGCAAAAATTATGAAGAAAGCCAGACACGAAACACACATAATAGC 1021
 Db 231 CysLysCysLeuAspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHis 250
 QY 1022 ATGCCAAAC-----TACATGCAG-----ACTGGAGGT 1048
 Db 251 GlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 270
 QY 1049 TTGGNAGCAGCCATTACCTG-----CCACCAACGAATTTTCAGGA----- 1090
 Db 271 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgLysGlnGluLeu 290
 QY 1091 -----CTTCCAAGATTACGTACCGAT 1111
 Db 291 PhePheAsnGlnGlyThrLysAspSerSerThrHisArgLeuGlyGlnAlaAsnAsnGly 310
 QY 1112 AGGGGCGCTTCTCTCA----- 1126
 Db 311 ArgThrThrSerSerGlnThrGlySerArgAlaGlyGlyAsnAlaSerLeuGlyProSer 330
 QY 1127 -----TGCATCTCTGGGAGGTGGTGGAGGCCACA 1156
 Db 331 LysValValTyrLysSerLeuLeuAlaAsnIleIleLysProMetAspValLysAlaLeu 350
 QY 1157 TGGCGCTGCTGCTTGTCTCAGGAGAGAGCCGAGAACACACTGCTCCCAAGTGCCTG 1216
 Db 351 CysSerValLeuValAlaValAlaGlyGluAlaAlaLysThrLeuThrGluLysArgLeu 370
 QY 1217 GCAGAGCAG 1225
 Db 371 AlaAsnGln 373

RESULT 8

Q9SZD1 ID Q9SZD1 PRELIMINARY; PRT; 603 AA.

AC Q9SZD1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative transcription factor.
 GN F19B15.30 OR AT4G29000.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Bancroft I., Meves H.W., Mayer K.F.X., Lemcke K., Schueller C.;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Meves H.W., Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL078470; CAB43914.1; -
 DR EMBL; AL161574; CAB79658.1; -
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2.
 SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AFE818DE5 CRC64;

Alignment Scores:

Pred. No.: 2,52e-26 Length: 603
 Score: 358.00 Matches: 93

RESULT
Q9CAV1

```

QY 494 CCTCAGTATCAAGATCAAAATAATATCTACAGTCAGAT-----GTC 535
Db 304 IleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThrIleAsnLysAla 323
QY 536 CTTAAACCAATGACTGCTTTA-----GTAGGAGATTTTGGCCAGCA 577
Db 324 ProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAspGluIleLeuProIle 343
QY 578 TCAACAAATAATCTCATTACACACAACTTGAGGACCCCTTA---CCATCGGTA--- 631
Db 344 ProArgThrIleGlyLeu-----HisLeuAsnGlyPheValAsnProSerValSer 360
QY 632 -----GTCAACGGGTCTCTCTTCCCTCGCGATCA----- 661
Db 361 SerGlyArgLysLysLysLysLysLysAspGlyClnAlaPheProSerThrThrPheHis 380
QY 661 ----- 661.
Db 381 TyrAsnIleGluAspGluPheSerThrProValSerThrLysArgAspLeuValValPhe 400
QY 662 -----ACTCTTCAGGACCAACCAAAATAACTTTGGCTGGG----- 697
Db 401 SerAspValLysIleMetGluProProGluArgSerValGluGluCysPheAspGln 420
QY 698 -----TACTGTGACTGCTTTGCCAGTGGGGACTTTTGGCAACAAC 736
Db 421 LeuMetAlaMetGluAsnArgTyrCysGluCysPheSerAlaGlyLeuPheCysGlyGlu 440
QY 737 ---TGCAATTGTAATAATTGTCACAACTTGTCATCATGATATTTGAACGGTTTAAAGCC 793
Db 441 ProCysSerCysGlnAsnCysPheAsnLysProIleHisGluAspLeuValMetLysSer 460
QY 794 ATTAAGCGCATGCTTGTCAGAAATCCAGAAGCTTTCAGCCCAAAATTTGGAAGGCCAA 853
Db 461 ArgGluValIleLysAlaArgAsnProLeuAlaPheAlaProLysValValSerThrSer 480
QY 854 -----TTGGCAATGTCAAG-----CCCCAGCACAAAC 880
Db 481 AspThrValIleAspLeuTrpValGluAsnSerLysThrProAlaSerAlaArgHisLys 500
QY 881 AAGGTGCAACTGCAGGAGGTGAGCTGCTGAAGAAATTTACTGCGAGTGTCTAGAGCC 940
Db 501 ArgGlyCysAsnCysArgLysSerGlyCysSerLysLysTyrCysGluCysPheMetMet 520
QY 941 CAATATTGTTCTCTATTTCAAATGCAATGCTGGTTGCAAAAT 985
Db 521 GlyValGlyCysSerSerAsnCysArgCysMetGlyCysLysAsn 535

RESULT 10
Q9LW71 ID Q9LW71 PRELIMINARY; PRT; 356 AA.
AC Q9LW71
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genomic DNA, chromosome 3, pl clone: MSL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty pl and TAC

```

```

RT clones.";
RL DNA Res. 7:131-135(2000).
DR ENBL: AB012247; BAB02682.1; -.
DR InterPro: IPR005172; CXC.
DR InterPro: IPR000345; CytC_heme_bind.
DR Pfam: PF03638; CXC; 2.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOMN_1.
SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DBFF58 CRC64;

Alignment Scores: . Length: 356
Pred. No.: 6,47e-17 Matches: 56
Score: 262.50 Conservative: 17
Percent Similarity: 45.34% Mismatches: 47
Best Local Similarity: 34.78% Indels: 41
Query Match: 6.86% Gaps: 5
DB: 10

US-09-743-237-3 (1-2134) x Q9LW71 (1-356)
QY 698 TACTGTGACTGCTTGGCAGTGGGACTTTTGGCAACATGCAATTCTAATAATTGTTC 757
Db 79 TyrCysAspCysPheAlaSerGlyValValCysThrAspCysAspCysValAspCysHis 98
QY 758 AACAACTGTCATCATGATATTGAACGGTTTAAAGCCATTAAGGCATGCTTGTGTAGAAAT 817
Db 99 AsnAsnSerGluLysCysAspAlaArgGluAlaAlaMetValAsnValLeuGlyArgAsn 118
QY 818 CCAGAACGTTTCCAGCCCAAAATTTGGAGAGGCCAATTTGGCAATGTCAAGCCCCAGCAC 877
Db 119 ProAsnAlaPheSerGluLys-----AlaLeuGlySerLeuThrAspAsnGln 134
QY 878 -----AACAAAGGTGCAACTGCAGAGG 901
Db 135 CysLysAlaAlaProAspThrLysProGlyLeuLeuSerArgGlyCysLysArg 154
QY 902 TCAGGCTGCCTGAAGAATTTACTGCGAGTGTATGAGGCCCAAAATTTATGTTCTTATT 961
Db 155 ThrArgCysLeuLysLysTyrCysGluCysPheGlnAlaAsnLeuLeuCysSerAspAsn 174
QY 962 TGCNAATGCATTTGTTGCAAAATTTATCAAGAACGCCAGAACCAAGAACTAATGAGC 1021
Db 175 CysLysCysIleAsnCysLysAsnValSerGluAla----- 186
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTGGAGGCGAGCCATTACCTGCCACCAACGAAA 1081
Db 187 -----PheGlnProProAla--- 191
QY 1082 TTTTCA-----GGACTTCCAAGATTCAGTCACGATAGCGGCCTTCTCTATGCATCTCC 1135
Db 192 PheSerAlaHisAsnSerProGlnValTyrArgArgArgArgArgGluLeuThrGlu 211
QY 1136 TGG 1138
Db 212 Trp 212

RESULT 11
Q92S22 ID Q92S22 PRELIMINARY; PRT; 896 AA.
AC Q92S22
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine-rich polycomb-like protein.
GN CPPL.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
NCBI_TaxID=3847;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanich C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,

```

Phlataski-Maunsbach K., Marcker K.A., Jensen E.O.;
 "cPpl, a novel type DNA-binding protein involved in the expression of
 a soybean leghemoglobin c3 gene.";
 Proc. Natl. Acad. Sci. U.S.A. 97: 8163-8168(2000).
 EMBL: AJ010165; CAA09028.1; -.
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98D7 CRC64;


```

RL DNA Res. 7:131-135(2000).
DR EMBL: AB022223; BAB01253.1; -.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76276 MW; 34BBA0E450F6BCE1 CRC64;

Alignment Scores:
Pred. No.: 6e-15 Length: 695
Score: 243.50 Matches: 131
Percent Similarity: 34.79% Conservative: 68
Best Local Similarity: 22.90% Mismatches: 173
Query Match: 6.37% Indels: 200
DB: 10 Gaps: 23

US-09-743-237-3 (1-2134) x Q9LUI3 (1-695)
QY 65 CCCGGGATCCAGAGCTCAGCGC-----GCTGGAGGAGCTGCG 103
DB 203 ProAsnAspSerGluAlaPheArgCysLeuValAspLysIleSerSerSerGluArgArg 222
QY 104 GCTCTCGAGCGCCGCGAGCGCCGCGC-----CTGCAACGTCCACTTCTGTCCTC 154
DB 223 PheCysAlaGlyValIleSerThrLysArgProAspIleAsnLysAspIleProAlaAsn 242
QY 155 GCTGCTACCCGCGCACCGCGCGCGGCTGTTTCCCGC----- 193
DB 243 GlySerSerAsnGluAsnGlnProLeuAlaValLeuProThrAsnGluSerValPheAsn 262
QY 194 CTGGGGCGGCTGGTCTCGGAAG-----AGCCTCCACCCGCGGCGTCC 238
DB 263 LeuHisArgGlyGly-MetArgArgCysLeuAspPheGluMetProGlyLysArgly 282
QY 239 GCATGATCCCACTGAAATCAAGGTAAGCAGGTGTTACTACTACAAGTAATAATCCGNA 298
DB 282 slyasp-----IleValAspAspI 289
QY 299 GAAGCAACTTTCAGAATCTTCTGTCAGGAATCCTGTTGCAAGTTCCATGGTCCCGAG 358
DB 289 nGlnSerValCys-AspAsnAsnValAlaGlyLysSerSerSerCysValValProG 309
QY 359 GAACCTAGAGGATGCTC-----CTGCTGTTCTCTTTAGAAAGATTCCAAACCAATGGTG 412
DB 309 lYileGlyLeuHisLeuAsnAlaValAlaMetSer-AlaLysAspSerAsnIleSerVal 328
QY 413 ATATGCCAATTGAAGGGGCGCACAAATGCTATGTATAGACAATCTAGACAACAGAGNA 472
DB 329 IleHisGlyTySerIleSerGlyGluIleGlnLysSerPheSerGlySerThrPro 348
QY 473 CTAAAGCACTCCATTTGTTCTCAGTATCAAGATCAA---AATAATTATCTACAGTCA 529
DB 349 IleGlnSerGlnAspThrValGlnGluThrSerAspGlnAlaGluAsnGluProValGlu 368
QY 530 GATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTTCGCCAGCATCAACAAATTA 589
DB 369 GluValProLys-----AlaLeuVal-----PheProGluLeu 379
QY 590 AATCTCATACACAACAACTTGAAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTC 649
DB 380 AsnLeu----- 381
QY 650 CCCTCGGGATCAACTCTTCAGGACCACCAAAAATAACTTTGGCTGGG----- 697
DB 382 -----GlySerLeuLysLysLysMetArgLysSerGluGlnAlaGlyGluGlyGluSer 399
QY 698 -----TACTGTGACTGCTTTGCC 715
DB 400 CysLysArgCysAsnCysLysSerLysCysLeuLysLeuTy-CysGluCysPheAla 419
QY 716 ACTGGGGACTTTTGC-----AACAACTGCAATTTGTAATTTGTCGAACAACTTGATCAT 772
DB 420 AlaGlyValTyCysIleGluProCysSerCysIleAspCysPheAsnLysProIleHis 439
QY 773 GATATTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGAACCTTTCCAG 832

```

```

DB 440 GluGluThrValLeuAlaThrArgLysGlnIleGluSerArgAsnProLeuAlaPheAla 459
QY 833 CCAAAATTT-----GGGAAGGGCCAAATGGGC 859
DB 460 ProLysValIleArgAsnAlaAspSerIleMetGluAlaSerAspAlaSerLysThr 479
QY 860 AATGTCACAGCCCGCACACAAGGGTCAACTCAGGAGCTCAGGCTGCCTGAAGAT 919
DB 480 ProAlaSerAlaArgHisLysArgGlyCysAsnCysLysSerAsnCysMetLysLys 499
QY 920 TACTCGAGTGTATCAGGCCCAATATGTCTTCTTATTTGCAAAATGCAATTTGTTCC 979
DB 500 TyrCysGluCysTyrglnGlyGlyValGlyCysSerMetAsnCysArgCysGluGlyCys 519
QY 980 AAAAATTAT----- 988
DB 520 ThrAsnValPheGlyArgLysAspGlySerLeuLeuValIleMetGluSerLysLeuGlu 539
QY 989 -----GAAGAAGCCGAGAACGAAAGACACTAATCAGCATGCCAAACTACATGCACT 1042
DB 540 GluAsnGlnGluThrTyrglnLysArgIleAlaLysIleGlnHisAsnValGluValSer 559
QY 1043 GGAGGTTTGGAA----- 1054
DB 560 LysGluValGluGlnAsnProSerSerAspGlnProSerThrProLeuProTyArg 579
QY 1055 -----GGCAGCCATTACCTGCCACCAACGAAATTT--- 1084
DB 580 HisLeuValValHisGlnProPheLeuSerLysAsnArgLeuProProThrGlnPhePhe 599
QY 1085 -----TCAGGACTTCCAAGATTTCAGTCACGATAGCGGCTTCTCATGTCATCTCCTGG 1138
DB 600 LeuGlyThrGlySerSerPhe-----ArgLysProAsnSerAsp--- 613
QY 1139 GAGGTGGTGGAGCCACATCGCCTCGCTTGTCTCAGGAGAGAGAGCCGAGAAAGAA 1198
DB 614 -----LeuAlaGlnSerGlnAsnGluLysLys--- 622
QY 1199 CACTGCTCAAGTGGCTGGCAGAGCAGATGATCTCGGAGGAATTTGGAAGGTGCTTATCA 1258
DB 623 -----ProLeuGluThrValThrGluAspLysThrGluIleMetPro 636
QY 1259 CAGATTCTCCACACTGAGTTTAAATCTAAGGATTCAAAATGGAGTAGAGTATAAGTCT 1318
DB 637 GluIle----- 638
QY 1319 GAATGCATGTTGATTTTGTCTTAGTCTAGAAAATCTCTAGTTTAGAAAGGATGTTAGGGG 1378
DB 639 -----LeuLeuAsnSerProIleAlaAsnIle-----LysAlaIleSerPro 652
QY 1379 AACATGAGCTGGCTCTGCAGACAAACAGCTCCCTGTCATCCCTGGCCGACGAGAT 1438
DB 653 AsnSerLysArgValSerProProGlnProGlySerSerGluSer-----GlySer 669
QY 1439 TTACTCAGAGCTCTCTGAAGATGTGCAAC 1468
DB 670 IleLeuArg-----ArgArgGlyAsn 676

```

Search completed: April 21, 2003, 12:01:00
Job time : 156.5 secs

